

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

MIGHTY ATOMS

A programmable quantum computer based on five atomic qubits **PAGES 35 & 63**

HUMAN PERFORMANCE

WE HAVE THE TECHNOLOGY

Bionic athletes prepare for 'cyborg Olympics' in Zurich

PAGE 20

ELECTRONIC WASTE

OFFSHORED POLLUTION

A global solution to the flood of toxic e-waste

PAGE 23

ECONOMICS

MINTING TROUBLE

Joseph Stiglitz on that conflicted currency, the euro

PAGE 26

NATURE.COM/NATURE

4 August 2016 £10

Vol. 536, No. 7614



THIS WEEK

EDITORIALS

WORLD VIEW Appeal of Donald Trump should shake scientists **p.7**

SPORT Ketone drink helps cyclists to push their physiology **p.8**



SALMON Wild fish swimming into farmed genetic trouble **p.8**

Ease pressure on PIs

Britain's Research Excellence Framework assessment needs to boost support for over-burdened principal investigators — a point missed in a review of the process.

Last week, a committee led by economist Nicholas Stern published recommendations on how to improve the United Kingdom's Research Excellence Framework (REF). Rightly, they support the principle and much of the practice of this periodic assessment of the research strengths of UK universities — which drives the allocation of universities' core funding. Although many academics resent having to submit their achievements every few years, the review concludes that the REF's substantial costs are greatly outweighed by its benefits.

Importantly, the recommendations seek to mitigate distortions introduced when institutions attempt to game the REF, for example by claiming credit for papers written by staff before they joined (see *Nature* <http://doi.org/bm9x>; 2016). They also support efficient documentation of the societal impacts of academic research.

But the report has not gone far enough on behalf of the linchpins of research: principal investigators. It has not recognized the threats to good science that arise from the overwhelming pressures now being placed on them.

The Stern review has made one very positive recommendation in that direction: it suggests that heads of academic departments and institutes could take the heat off their staff by no longer requiring each investigator to submit the same number of research outputs to the REF. For the next assessment, in a few years' time, departments would instead be required to submit only a certain number of outputs overall: some principal investigators might report more than average; some might even report none. A good institute head would balance their virtues on the basis of the long-term character of their research.

The motivation for this is clear when the review says that a priority

is to find “ways to ensure that the REF can encourage researchers to explore big or fundamental problems, in ways that may not deliver a steady stream of papers or a quick monograph; to deliver academically excellent synthesis of evidence and meta-analysis to support policy making; and to develop game changing ideas that, for example, can lead to the development of new disciplines, or that have significant impact outside their discipline”.

But the pressures on principal investigators arise not only from research accountability. Alas, researchers are merely human. They have finite bandwidths, and it is difficult to balance their duties when journals, funders and universities are rightly increasing their demands for better data management and sharing, better reproducibility, better mentoring of postdocs and graduate students, better teaching and broader stakeholder engagement. No wonder many of the best principal investigators are wilting under the stress, and even leaving academia. This is exacerbated when funding allocation is ultracompetitive.

The REF should attend to this when it assesses a department's research environment. Stern's recommendations would empower universities to strengthen research cultures. But institutions and funders should act directly to mitigate pressures on principal investigators, for example by supporting staff for data-management planning and sharing, crafting grant applications and administrative tasks. This would combat a creed that research money is best spent only on yet more postdocs.

The REF should help institutions to counter such instincts without compromising the creative autonomy of the principal investigators on whom they depend. The Stern review could and should have pushed harder in that direction. ■

Cyborg Olympics

The Olympic Games celebrate physical skill; the Cybathlon honours innovation in prosthetics.

What defines human physical excellence? Is it the pain, sweat and grit of elite athletes using every slight genetic advantage to perfect their bodies for competition? Or is human ingenuity also to be celebrated, particularly when science can allow disabled athletes — who are just as gritty and driven as their able-bodied counterparts — to compete on a level playing field?

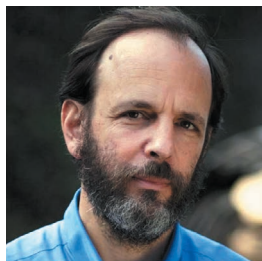
The Olympics (see page 18) and Paralympics already struggle with this question. Now, into the debate comes a ‘cyborg Olympics’ that melds human and machine to create a new kind of athlete. In October, nearly 80 teams from 25 countries around the world will gather in Zurich, Switzerland, to compete in the Cybathlon (see page 20).

Each team is made up of engineers and scientists who have created a powered prosthesis for a disabled ‘pilot’ to use in one of six competitions. Electrical stimulation of paralysed leg muscles allows pilots with spinal-cord injuries to ride bikes. Other races use robotic prosthetic arms to complete tasks such as setting a dinner table, or track brain activity to race avatars on a screen.

What sets the Cybathlon apart from other sporting events is how it celebrates human technological achievement rather than just physical excellence. The Olympic Games strictly limit the technology that athletes can use, for instance requiring cyclists to ride bikes that adhere to tight standards. The Cybathlon, by contrast, limits the humans, requiring that its cyclists must not be able to move their legs without the help of artificial stimulation.

The goals of the two events are very different, of course. The Olympics is a competition for fans' entertainment and athletes' glory, whereas the Cybathlon is intended to kick-start innovation in prosthetics for real-world uses. And as technology and opportunities develop, they should also spark broader debate about human enhancement. ■

MICHAEL TEMCHINE



Donald Trump's appeal should be a call to arms

Trump's nomination as Republican presidential candidate is a reminder that scientific progress has not benefited all Americans, says **Daniel Sarewitz**.

If, as the French counter-revolutionary Joseph de Maistre wrote in 1811, every nation gets the government it deserves, what might the United States have done to deserve Donald Trump?

A well-functioning democracy should undercut the appeal of blustering, xenophobic demagogues by ensuring that most citizens have a stake in government and hope for the future. And although no single cause or problem can explain Trump's appeal to a large part of the American electorate, his nomination as the Republican presidential candidate should be cause for serious reflection about what is going wrong in America. For many Americans, one thing that has gone wrong is that the promise of scientific and technological progress has not been fulfilled.

This promise is at the heart of the American identity: it is anchored by founding fathers Benjamin Franklin and Thomas Jefferson, scientists and inventors both, extolled by Alexis de Tocqueville in his 1835 masterwork *Democracy in America*, embodied in the inventions of Thomas Edison, and codified in its modern form in *Science, The Endless Frontier*, Vannevar Bush's famous 1945 science-policy report to President Franklin D. Roosevelt, which laid out the still-powerful argument for government sponsorship of basic science.

Indeed, Bush's linking of the frontier metaphor to the promise of scientific progress was a distinctively American flourish. And his formula was simple: three factors — “the free play of initiative of a vigorous people under democracy, the heritage of great natural wealth, and the advance of science and its application” — would deliver to all Americans full employment and rising standards of living, improved health, and military security. Government investment in science, especially research carried out at the nation's elite universities, would prime the pump of continual progress.

Not everyone, however, was buying Bush's story. Starting in the early 1940s, Senator Harley Kilgore, a Democrat from West Virginia, championed a different national approach to science policy, one in which government investment would focus research and development directly on social goals and economic growth. A six-year political battle between Kilgore and Bush followed, to control not just US science policy itself, but, equally importantly, the rhetoric of science and progress. Bush, who had much of the leadership of academic and industrial science on his side, and who saw Kilgore as a threat to the independence of both elite academic science and the economic marketplace, became the decisive winner on both fronts: the 1950 bill creating the National Science Foundation gave scientists primary responsibility for determining the agency's research agenda.

Over the subsequent 65 years, scientists and science advocates have not shirked from parroting Bush's *Endless Frontier* vision of

scientific knowledge, flowing from “the free play of free intellects”, as an unalloyed good from which all citizens would benefit through the ever-expanding economic opportunities created by science-based innovation. It has been an appealingly non-ideological view of progress, adopted across the political spectrum. As Nobel-prizewinning physicist Leon Lederman put it in 1992: “What's good for American science ... is good for America.”

Maybe not. Although Trump supporters are by no means a homogeneous lot, a clever analysis in *The New York Times* in March showed that they can most reliably be characterized by two attributes. First, they identify their ancestral heritage as American, rather than any particular ethnic or religious stock. And second, they live in regions of the country that have not only failed to benefit economically from innovation, but have been harmed by it.

Mainstream media analysis of the Trump phenomenon almost never links it to the science and technology policies pursued by the nation since the Second World War. Yet technological revolutions arising from these policies have contributed to more than 40 years of wealth inequality, disappearing middle-class jobs and eviscerated manufacturing communities in the places where support for Trump is strongest. Indeed, economic theory throws aside these millions of people as the inevitable losers in the ‘creative destruction’ that science catalyses, as if ruined cities and livelihoods are just side effects of the strong medicine of science-based innovation. These people are the cost of the prevailing myth of progress, and, given their core identity as ‘Americans’, it is no wonder

they are susceptible to Trump's jingoistic populism.

No one remembers Harley Kilgore any more, and it's impossible to know whether his socially oriented vision of science policy might have contributed to a more equitable linking between scientific advance and economic benefit. But it is more than simply ironic that Kilgore's home state of West Virginia — whose per capita income ranks 49th out of the 50 states — is now Trump's strongest supporter.

Having claimed for more than a half a century that science-based innovation would be good for everyone, science advocates and scientists who have benefited so greatly from this line of argument can hardly now say, “Oh, but it's not our fault, these are problems of trade and labour and economic policy”. Trump's ascendance should rekindle the Bush–Kilgore debates, and policymakers should seriously consider what a system of socially responsible and responsive science would look like. The current system has failed the test. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: daniel.sarewitz@asu.edu

**THE PROMISE
OF SCIENTIFIC AND
TECHNOLOGICAL
PROGRESS
HAS NOT BEEN
FULFILLED.**

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

MICROSCOPY

Single ions make sharper images

A microscope creates images with nanometre resolution by exposing samples to single ions.

In electron and ion microscopy, increasing a sample's exposure time can improve the signal-to-noise ratio and result in clearer images, but this can damage or contaminate the sample. To avoid this, Georg Jacob at the University of Mainz in Germany and his colleagues used an electric-field 'trap' to release calcium ions one by one.

Each ion, either transmitted or blocked by the sample, corresponds to a pixel. By controlling the release of the ions, the team calculated when those coming from the source and through the sample should arrive at the detector. This allowed the team to turn the detector on only during those times, reducing the number of detected 'noise' ions.

The microscope showed a fivefold increase in the signal-to-noise ratio, and could potentially cut noise signals by a factor of one million, compared with current methods. It also pinpointed the position of a 1-micrometre hole in a diamond sample with a precision of 2.7 nanometres. *Phys. Rev. Lett.* 117, 043001 (2016)

EXERCISE PHYSIOLOGY

Ketones alter metabolism

Athletes' physical endurance can be enhanced by drinking ketones — biochemical fuel normally produced during starvation.

Fasting or prolonged exercise drives liver cells to make ketone bodies as

a fast-acting fuel to help tissues cope with the energy deficit. To test the effect of these molecules on exercise metabolism, Pete Cox at the University of Oxford, UK, and his colleagues gave endurance athletes a drink containing ketones. They found that after a prolonged period of cycling, the metabolism of those who consumed ketones had shifted so that they conserved glucose and burned more fat than those who did not receive ketones. In a 30-minute time trial done after one hour of

cycling, athletes who had consumed ketones and carbohydrates cycled more than 400 metres further, on average, than those who had eaten only carbohydrates. *Cell Metab.* <http://doi.org/bm8z> (2016)

ANIMAL BEHAVIOUR

Humpbacks to the rescue

Humpback whales come to the assistance of other species by harassing the killer whales

that are attacking those animals.

Robert Pitman at the Southwest Fisheries Science Center in La Jolla, California, and his colleagues reviewed reports of 115 interactions between humpback whales (*Megaptera novaeangliae*) and killer whales (*Orcinus orca*). In at least 31 cases, humpbacks approached and 'mobbed' killers as the killers attacked or fed — mostly on sea lions, seals and other whale species.

The authors suggest that the



NPL/ALAMY

CONSERVATION

Farmed salmon go wild

Norway's wild salmon owe part of their genetic make-up to escapees from salmon farms, which could compromise the fitness of the wild population.

Wild Atlantic salmon (*Salmo salar*; pictured) are more genetically diverse and generally better adapted to the environment than are their farmed counterparts. Sten Karlsson and his colleagues at the Norwegian Institute for Nature Research in Trondheim looked at genetic markers in 21,562 wild salmon from 147 locations around Norway. They found

significant genetic material from farmed salmon in wild fish from about half of these locations, with the average wild population showing 6% farmed genetic heritage. In some locations, this rose to 42%.

Wild populations in areas with many salmon farms contained higher levels of farmed salmon DNA than did those in regions with less farming. Managers of both wild and farmed animals should work to minimize mating between the two populations, the authors say. *ICES J. Mar. Sci.* <http://doi.org/bm6j> (2016)

humpbacks responded to the calls of the attacking killers. They add that humpbacks do not seem to benefit from helping other animals, so this may be a case of interspecies altruism.

Marine Mammal Sci. <http://doi.org/bm58> (2016)

NEUROSCIENCE

Fruit flies care about texture

Fruit flies prefer food with certain textures, thanks to specific neurons in the brain that connect to taste sensors in the tongue.

Yali Zhang and Craig Montell at the University of California, Santa Barbara, and their colleagues gave fruit flies (*Drosophila melanogaster*) liquid solutions of varying viscosity, and solid foods of varying hardness. The flies preferred food that was less viscous and of an intermediate hardness. The authors identified a type of neuron, called md-L, that responded to mechanical stimulation of the sensory hairs on the tongue. When the team stimulated these neurons, the flies' feeding behaviour altered according to the strength of the stimulus.

The researchers pinpointed a protein, TMC, that is found in the neurons' membranes and is required for texture sensation. Mammals and other animals also express TMC proteins, suggesting that they may also have neuronal sensors for food texture.

Neuron <http://doi.org/bm8x> (2016)

IMMUNOLOGY

Gut microbes boost antibodies

Intestinal bacteria release metabolic by-products that support antibody-producing immune cells.

Gut microbes produce short-chain fatty acids as they digest dietary fibre. Chang Kim and his colleagues at Purdue University in West

Lafayette, Indiana, treated cultured B cells with the fatty acids and found that this enhanced the expression of genes that help the cells to develop into antibody-producing factories known as plasma B cells. The treatment also increased the cells' metabolism, helping to support the energy-consuming process of making antibodies.

Mice fed a low-fibre diet were more susceptible than other animals to infection by the pathogen *Citrobacter rodentium* and had weaker immune responses. Treating the mice with short-chain fatty acids or dietary fibre increased antibody production and reversed this immune deficiency.

Cell Host Microbe <http://doi.org/bm82> (2016)

EVOLUTION

Long trips foster tool use

Chimpanzees that travel farther in search of food are more likely to use tools than are their less-travelled counterparts.

Thibaud Gruber and his colleagues at the University of Neuchâtel in Switzerland studied wild chimpanzees (*Pan troglodytes schweinfurthii*; pictured) in Uganda, using six years of experimental data and seven years of observations. The researchers drilled small holes in logs and filled them with honey, which the chimps



could access only using a tool such as a leaf sponge (pictured). Chimps were more likely to use tools to get honey when they had recently foraged over longer distances, compared to those that travelled less. Published data on wild chimps also revealed that communities that travel more use a larger repertoire of feeding tools.

Moving long distances may have helped to drive the development of early human technology, the authors say. *eLife* <http://doi.org/bm6c> (2016)

ENVIRONMENTAL SCIENCE

Humans have a hand in wildfires

People strongly influence the likelihood of fires in forests, grasslands and other ecosystems across the United States and Canada, mostly by lowering fire risks.

People can ignite fires, but can also suppress them by altering properties of the land, for example by removing natural vegetation. To better understand the effects humans have, a team led by Marc-André Parisien of Natural Resources Canada in Edmonton used statistical models to analyse human and natural factors linked to fire probability across both countries between 1984 and 2014. They found an association between human activities and fire — the stronger the human influence, the lower the probability of fire.

Wildfires (pictured) are rarely purely natural, and fire managers should take this into account when considering how fire risks may shift in a warming world, the authors say.

Environ. Res. Lett. 11, 075005 (2016)

ECOLOGY

Insecticides hurt male bees too

A class of pesticides that has been linked to declining bee populations harms the reproductive capacity of male honeybees, not just that of queens as other research has shown.

Neonicotinoid pesticides are currently banned by the European Union because of their effects on bees. Lars Straub at the University of Berne and his team exposed colonies of honeybees (*Apis mellifera*) in the field to two neonicotinoids that are commonly found in agricultural fields. The team found that in males, the chemicals reduced living sperm count by 39% and decreased the insects' lifespan.

This could negatively affect honeybee colony fitness and queen survival and health, the authors say.

Proc. R. Soc. B 283, 20160506 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Hinkley deal delay

In a surprise move, the UK government delayed its final approval of a new nuclear power plant at Hinkley Point on 28 July, just hours after French energy company EDF, which is financing most of the project's construction, gave it the go-ahead. Hinkley Point C would be the first new UK nuclear power station this century, and has an £18-billion (US\$24-billion) price tag. China has signed up to provide one-third of the cost. The project was championed by the previous UK government, but the current government said that it needed time to review the deal.

Farewell Philae

On 27 July, the European Space Agency (ESA) switched off radio communications with Philae, the space probe that made history by landing on a comet in November 2014. Philae had a bright but unlucky career. After landing on the comet 67P/Churyumov–Gerasimenko, it failed to grip the comet's surface and

SOUND BITE

“I believe in science. I believe that climate change is real.”

Hillary Clinton gives time to science in her speech accepting the Democratic nomination for US president at the party's convention in Philadelphia, Pennsylvania, on 28 July.



YANNIS BEHRAKIS/REUTERS

EU funds research on migrant crisis

The European Commission is making €11 million (US\$12.3 million) available for research that addresses challenges related to migration. Some 1.25 million refugees have entered Europe since the start of 2015, but management policies across the continent are weak and poorly coordinated. As part of

its Horizon 2020 framework programme, the commission will next year announce five calls for proposals related to different policy areas, including the integration of migrants into the workforce and society. Carlos Moedas, European Commissioner for Research, Science and Innovation, announced the measures last week.

bounced into a shady spot where it was unable to charge its solar panels. It performed just 64 hours of experiments before its batteries died. The probe signalled again briefly in June and July 2015, but by February, ESA said that there was “close to zero” chance of hearing from it again. Philae's parent satellite, Rosetta, is preparing for its own demise. It will crash land on 67P on 30 September, and scientists hope that it will collect a rich trove of cometary data on its approach.

Zika in Florida

Fourteen cases of Zika virus in Miami, Florida, are likely to have arisen from local *Aedes aegypti* mosquitoes, state authorities said on 1 August. Although this would be the

first time that people in the continental United States have become infected at home and not abroad, US health authorities have long predicted that such small local clusters would be seen in some US southern states — but not elsewhere in the country. On the basis of past experiences with dengue and chikungunya, viruses spread by the same mosquito, US authorities do not expect to see any widespread transmission, in sharp contrast to the situation in most of the 50 countries and territories that since 2015 have reported their first ever outbreaks of Zika.

Audit revisions

A government-commissioned report on the Research Excellence Framework (REF),

the United Kingdom's periodic audit of the quality of its research, has called for changes to the system in a bid to cut costs and prevent ‘gaming’. The REF, which occurs every 5–7 years, is used to allocate about £2 billion (US\$2.6 billion) in annual research funding. Among its suggestions, the report proposed that the audit should not give universities credit for papers written by staff before they joined, and that institutions should submit all their researchers to the audit process, rather than just a selection. The government will consider the recommendations in the report, which was led by economist Nicholas Stern and released on 28 July, before issuing a formal response. See page 5 and go.nature.com/2ardiyc for more.

RESEARCH

New whale species

Scientists have identified a new species of whale, after a DNA analysis of 178 beaked whales revealed a genetically distinct subset. The species, in the genus *Berardius*, is found in the Okhotsk and Bering seas in the northern Pacific and was previously considered to be a dwarf form of Baird's beaked whale (*B. bairdii*). Japanese whalers had long acknowledged two distinct variants of the whale: the common 'slate-grey' variety, and the newly recognized 'black' form, which they called *karasu*, Japanese for raven. The discovery was published on 26 July (P. A. Morin *et al. Mar. Mamm. Sci.* <http://doi.org/bm7b>; 2016).

PEOPLE

NIMH chief

Psychiatrist and neuroscientist Joshua Gordon will be the next head of the US National Institute of Mental Health (NIMH), it was announced on 28 July. Gordon (**pictured**) will join the institute from the Columbia University Medical Center in New York City, where his research has focused on schizophrenia and anxiety, and on how genetic mutations lead to particular behaviours. Gordon is



expected to take up the post at the NIMH, which has an annual budget of about US\$1.5 billion, in September.

BUSINESS

GSK backs Britain

Drug giant GlaxoSmithKline (GSK) said on 27 July that it will invest £275 million (US\$363 million) into its UK operations, allaying fears that big pharmaceutical companies would begin to move out of the United Kingdom after the country's vote last month to leave the European Union. GSK will put the money into three manufacturing sites, to support production of respiratory and large-molecule biological medicines.

Indian space case

The Indian Space Research Organisation (ISRO) received a major blow on 25 July, when the Permanent Court of Arbitration in The Hague, the Netherlands, ruled against

it in a contract dispute. In 2011, ISRO's commercial arm, Antrix Corporation, scrapped a 2005 deal with satellite company Devas Multimedia in Bangalore for the long-term lease of two satellites, citing allegations of irregularities in the deal, including security concerns. Devas took the matter to international arbitration courts. Industry observers say that the Indian government may have to pay about US\$1 billion in damages to Devas.

Bioelectronics firm

London-based drug giant GlaxoSmithKline announced on 1 August that it is teaming up with Verily, Google's life-sciences spin-off in South San Francisco, California, to develop electronics-based therapies. The two companies will contribute up to £540 million (US\$713 million) to the newly established Galvani Bioelectronics, which will be based in the United Kingdom, over 7 years. Galvani will develop miniature implants that can alter the body's electric nerve signals, with the aim of treating inflammatory, metabolic and hormone conditions.

Theranos device

Elizabeth Holmes, chief executive of the beleaguered biotechnology firm Theranos, unveiled a new blood-testing

COMING UP

11–12 AUGUST

Chemistry graduates gather in Oxford, UK, for the Oxford Synthesis Summer Conference.

go.nature.com/2aunyt

12 AUGUST

The annual Perseids meteor shower reaches its peak.

machine on 1 August at a meeting of the American Association for Clinical Chemistry in Philadelphia, Pennsylvania. Theranos rose to prominence with promises of a technology that could perform a wide range of diagnostic tests using a few drops of blood. But its claims faced scepticism and government scrutiny, and in July, US regulators banned Holmes from running a lab for two years. Theranos, of Palo Alto, California, says that its new miniLab machine can perform a variety of tests from a finger-prick of blood, but the device has not been independently verified.

POLICY

Inclusive astronomy

On 28 July, the American Astronomical Society endorsed a statement intended to improve the experience in astronomy of under-represented groups including women, ethnic and racial minorities, disabled people and some sexual and gender identities. Suggested changes include eliminating discriminatory hiring practices and ensuring that astronomy institutions, facilities and data are accessible to everyone. The recommendations, spurred in part by discussions about inequality in science, came from a meeting on inclusive astronomy held in Nashville, Tennessee, in June 2015.

NATURE.COM

For daily news updates see:

www.nature.com/news

TREND WATCH

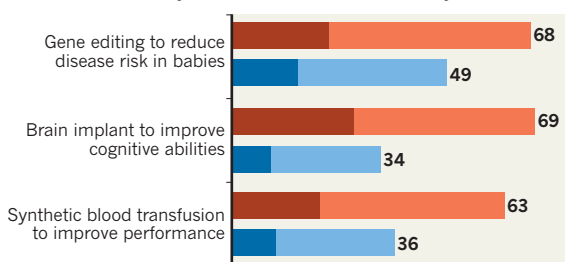
More than 60% of people in a US survey are concerned about scientific advances being used for human 'enhancement', revealed the Pew Research Center in Washington DC on 26 July. The poll asked 4,726 people how they felt about three potential technologies: gene editing to reduce disease risk in babies; brain implants to enhance brain processes; and transfusions of synthetic blood to improve strength. In each case, fewer than half expressed enthusiasm. See go.nature.com/2ag0jjd for more.

US PUBLIC WARY OF HUMAN ENHANCEMENT

Most people in the United States are concerned about the prospect of using cutting-edge science to enhance human abilities.

How (a) worried and (b) enthusiastic do you feel about each of these ideas?

Enthusiastic: ■ Very ■ Somewhat Worried: ■ Very ■ Somewhat



Percentage of respondents

Based on survey answers from 4,726 US adults in March 2016.

NEWS IN FOCUS

MATHEMATICS Monumental proof continues to torment number theorists **p.14**

POLITICS British scientists seek influence in new 'Brexit ministry' **p.15**

SCHOLARLY OLYMPICS A graphical guide to the games' impact on research **p.18**



PROSTHETICS Cyborg athletes tackle the everyday world **p.20**

ASAHI SHIMBUN/GETTY



The Hitomi X-ray astronomy satellite launched in February, but broke up in space after a month.

ASTRONOMY

Troubled Japanese space agency seeks fresh start

Push to resurrect instrument lost during satellite failure highlights JAXA's resilience.

BY ALEXANDRA WITZE

The Japan Aerospace Exploration Agency (JAXA) is on a quest for redemption. In March, a software error caused the agency's Hitomi X-ray astronomy satellite to break up in space, cutting short a planned three-year mission after only one month.

Now JAXA is considering whether to rebuild and relaunch a copy of the spacecraft's key instrument — a US-built X-ray spectrometer — with help from NASA. On

5 August, representatives of the two space agencies will meet to discuss the possibility of resurrecting the instrument that was the heart of Hitomi's science. But whether JAXA can regain the confidence of the Japanese nation, and of its international partners, remains to be seen.

Space experts note that JAXA has pulled off stunning recoveries before. It coaxed its crippled Hayabusa spacecraft to bring back dust from an asteroid, and nudged its Akatsuki probe into orbit around Venus 5 years

after an engine failure seemed to render the spacecraft useless.

"It's important to note how resourceful JAXA has been at recovering from failures that typically would be catastrophic," says Ralph Lorenz, a planetary scientist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland, and co-author of the book *Space System Failures* (Praxis, 2005).

Hitomi broke apart because an erroneous software command prompted the spacecraft to spin faster and faster, until its ►

► solar panels flew off into space. A JAXA investigation blamed faulty project-management techniques for not catching the error.

The failure has reverberated at every level of JAXA's Institute of Space and Astronautical Science (ISAS) in Sagami-hara, which managed Hitomi. JAXA president Naoki Okumura was one of three leading officials who took a 10% pay cut for four months "to express our regret and caution ourselves", he said in a June press conference. He has also ordered a systems review of the institute's next big project: a mission to study Earth's radiation belts that is slated to launch in the coming months.

Before Hitomi, JAXA's lowest point was perhaps the loss of its Nozomi mission to Mars, which sailed past the red planet in 2003 without entering orbit as it was supposed to. The same year, a new JAXA rocket design failed during a test launch, prompting a review of all agency projects.

TRY, TRY AGAIN

Some have questioned whether JAXA is trying to do too much with too little. It often assigns one person to cover a number of tasks that NASA would spread among multiple project engineers, says Lorenz, who collaborates on the Akatsuki Venus probe.

Okumura has acknowledged as much, saying that ISAS will generally develop a mission using a small in-house team, along with the

spacecraft manufacturer. By contrast, Hitomi involved a larger number of complex systems. There were simply not enough safeguards built into the process to catch the software error. "The previously conventional ISAS methods were not necessarily suited for the production

"It's important to note how resourceful JAXA has been at recovering from failures."

of modern satellites and spacecraft," Okumura said.

JAXA has released an extraordinary level of technical detail about the failure. Agency officials have said that

because Hitomi was meant as a community mission to serve X-ray astronomers across the globe, they feel obligated to explain what happened so that nobody makes the same mistake.

Because of this determination and openness, "I think Hitomi's successor is in safe hands with JAXA," says Elizabeth Tasker, an astrophysicist at Hokkaido University in Sapporo, Japan.

But such projects may be a hard sell to politicians. "High-profile setbacks like Nozomi and Hitomi make it difficult for JAXA to justify big-ticket science missions in today's political atmosphere," says Saadia Pekkanen, an expert in Japanese space policy at the University of Washington in Seattle.

JAXA has not yet decided whether a Hitomi

successor would fly or which instruments it would carry, says ISAS spokeswoman Chisato Ikuta. But Hitomi's premier scientific instrument was the spectrometer provided by NASA; data that it collected before the spacecraft died revealed secrets about gas flows in the Perseus galaxy cluster.

The spectrometer seems to be thrice cursed; two earlier versions on different satellites were lost to a launch failure and a coolant leakage. Even so, a NASA advisory group reported on 5 July that launching a copy of the instrument no later than 2023 "would fulfill the immense scientific promise of the Hitomi" spectrometer. The cost to rebuild would be roughly US\$70 million to \$90 million.

Paul Hertz, NASA's astrophysics director, will meet with JAXA representatives to discuss the options. "Certainly we would not be overseeing JAXA," he told a NASA advisory committee on 20 July. "We can discuss practices that NASA implements to prevent us from making avoidable mistakes."

Other international missions in the works from JAXA include a magnetospheric orbiter, which is scheduled to launch next year on the European Space Agency's BepiColumbo mission to Mercury.

"The Olympics of engineering is when things go wrong," says Lorenz. "Maybe the best time to fly is right after a failure." ■

MATHEMATICS

Grand proof fazes theorists

Conference on Shinichi Mochizuki's 'revolutionary' work inspires cautious optimism.

BY DAVIDE CASTELVECCHI

Nearly four years after Shinichi Mochizuki unveiled an imposing set of papers that could revolutionize the theory of numbers, other mathematicians have yet to understand his work — although they have made modest progress.

Some four dozen mathematicians converged in Japan last week for a rare opportunity to hear Mochizuki present his monumental proof of the 31-year-old *abc* conjecture, which sits at the heart of number theory. The conference took place on his home turf, at Kyoto University's Research Institute for Mathematical Sciences (RIMS).

Mochizuki is "less isolated than he was before the process got started", says Kiran Kedlaya, a number theorist at the University of California, San Diego. At first, Mochizuki's proof, which stretches over more than 500 pages (available at go.nature.com/2amidei), seemed like an

impenetrable jungle of formulae. But experts have slowly discerned a strategy, and have zeroed in on particular passages that seem crucial, Kedlaya says.

And Jeffrey Lagarias, a number theorist at the University of Michigan in Ann Arbor, says that he got far enough to see that Mochizuki's work is worth the effort. "It has some revolutionary new ideas," he says.

Still, Kedlaya says that the more he delves into the proof, the longer he thinks it will take to reach a consensus on whether it is correct. He used to think that the issue would be resolved perhaps by 2017. "Now I'm thinking at least three years from now."

Others are even less optimistic. "The constructions are generally clear, and many of the arguments could be followed to some extent, but the overarching strategy remains totally elusive for me," says mathematician Vesselin Dimitrov of Yale University in New Haven, Connecticut. "Add to this the heavy,

unprecedentedly indigestible notation: these papers are unlike anything that has ever appeared in the mathematical literature."

THE *ABC* PROOF

The *abc* conjecture relates to prime numbers — whole numbers that cannot be evenly divided by any smaller number except 1. The conjecture comes in a number of different forms, and explains how the primes that divide two numbers, *a* and *b*, are related to those that divide their sum, *c*.

If Mochizuki's proof is correct, it would have repercussions across the entire field, says Dimitrov. "When you work in number theory, you cannot ignore the *abc* conjecture," he says. "This is why all number theorists eagerly wanted to know about Mochizuki's approach." For example, Dimitrov showed in January how, assuming the correctness of Mochizuki's proof, one might be able to derive many other results, including an independent proof

of the celebrated Fermat's last theorem (V. Dimitrov. Preprint available at <http://arxiv.org/abs/1601.03572>; 2016).

The purported proof, which Mochizuki first posted on his webpage in August 2012, builds on more than a decade of previous work, in which he developed a novel and extremely abstract branch of mathematics in virtual isolation.

MOCHIZUKI IN THE ROOM

The Kyoto workshop followed on the heels of one held last December in Oxford, UK. Mochizuki did not attend that first meeting, although he answered questions over a video link. This time, having him in the room — and hearing him present some of the materials himself — was helpful, says Taylor Dupuy, a mathematician at the Hebrew University of Jerusalem.

Around ten mathematicians are now putting substantial effort into digesting the material — up from three before the Oxford workshop, says Ivan Fesenko, a mathematician at the University of Nottingham, UK, who co-organized both workshops.

Mochizuki did not take part in the customary mingling and social activities at the Kyoto meeting. And although he was unfailingly forthcoming in answering questions, it was unclear what he thought of the proceedings. "Mochizuki does not give a lot away," Kedlaya says. "He's an excellent poker player."

Mathematicians have criticized Mochizuki for his refusal to travel: after he posted his papers, he turned down multiple offers to go abroad. He spent much of his youth in the United States, but is now said to rarely leave the Kyoto area. (Mochizuki does not respond to interview requests, and the workshop's website noted: "Activities aimed at interviewing or media coverage of any sort within the facilities of RIMS, Kyoto University, will not be accepted.")

"He is very level-headed," says another workshop participant who did not want to be named. "The only thing that frustrates him is people making rash judgemental comments without understanding any details." Still, Dupuy says, "I think he does take a lot of the criticism about him really personally. I'm sure he's sick of this whole thing, too." ■



MATTHEW CHATFIELD/REX/SHUTTERSTOCK

David Davis leads the UK government's Department for Exiting the European Union.

POLITICS

UK scientists seek Brexit influence

They hope for active role in negotiations to exit EU.

BY ELIZABETH GIBNEY

British science's largest lobbying campaign in years is under way. After the shock of the United Kingdom's vote to leave the European Union, anxious researchers are doing all they can to ensure that their interests are represented in Brexit negotiations. One big unanswered question is what role science will have in the new 'Brexit ministry' — the Department for Exiting the European Union (DEEU) — that has been expressly formed to take the country out of the EU.

Worried at the prospect of losing access to EU funding and collaborations, scientific societies have fired off numerous letters asking the government to keep their country in the EU's research system, and warning of damage already caused by Brexit. An advocacy group, Scientists for EU, says it has gathered (in confidence) 25 cases of foreign scientists withdrawing job applications or being refused a UK

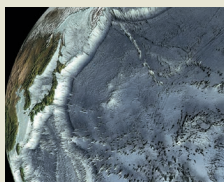
post as a result of Brexit, 7 cases of someone in UK science leaving the country, and 33 of disruption to funding for the EU's Horizon 2020 research-grants programme.

The government has indicated that it is listening to scientists — but seems reluctant to say so too loudly. On 18 July, Prime Minister Theresa May sent a letter to Paul Nurse, the director of London's Francis Crick Institute, telling him that the government was committed to "ensuring a positive outcome for UK science" as the country exited the EU. But the letter — effectively May's first statement on science — did not become public knowledge until science minister Jo Johnson referred to it in passing in a 25 July speech at the EuroScience Open Forum in Manchester, prompting journalists to press for a copy. Venkatraman Ramakrishnan, the president of London's Royal Society, said he welcomed the comments and was looking forward to working with May and her colleagues "to turn these words into action". ▶



**MORE
ONLINE**

TOP NEWS



The fiery birth of Earth's largest ocean
go.nature.com/2aj7q8g

MORE NEWS

- Philae comet lander goes quiet for good go.nature.com/2az8klh
- Major review calls time on 'gaming' in UK research assessment go.nature.com/2ardiyc
- Women in physics face big hurdles — still go.nature.com/2aor9lt

NATURE PODCAST



How thirst works, a programmable quantum computer, and how children learn
nature.com/nature/podcast

► What action May will take remains unclear: prospects for science are inextricably entangled with the wider Brexit issues of freedom of movement and UK access to the EU's single market. David Davis, a Member of Parliament who had campaigned on the 'leave' side of the referendum, leads the DEEU. He has announced plans to conduct a "huge consultation" ahead of the start of formal EU exit negotiations, which May has postponed until at least 2017.

SCIENCE IN THE BREXIT MINISTRY

Davis's team is talking to "the research institutes", he told Sky News on 17 July — but his department could not confirm which bodies this referred to. UK national academies have written jointly to Davis and "look forward to working with him to ensure that science's voice is heard in Brexit negotiations", the Royal Society told *Nature*.

Some hope that the Brexit ministry will contain specific advocates for research. "There should be some sort of champion for science within the department," says John Beddington, a population biologist at the Oxford Martin School, and a former UK chief scientific adviser. An obvious choice is science minister Johnson, Beddington says, although the DEEU could also dedicate a group of civil servants to the job. Johnson could be a "very strong, very early voice" in DEEU deliberations, Sharon Witherspoon, policy chief at the UK Academy of Social Sciences, told a House of Lords inquiry on 19 July. She added that research needed "urgent attention, and cannot wait to be an afterthought".

Giving more-formal responsibilities to Johnson, whose role in May's government is split between the education and business departments, might be a stretch. "If anyone can do it, Jo can. But I'm not confident that the best voice for the science community would be to add another job on for Jo," says Nick Hillman, director of the Oxford-based Higher Education Policy Institute.

A different potential conduit for scientific input could be the DEEU's departmental board, an advisory body that, in other departments, often includes senior business figures. And another idea is for Davis's department to appoint a chief scientific adviser (CSA), as most other UK ministries already have. But Beddington says that although the DEEU and the newly created Department for International Trade should each have a CSA, their role should not be to advocate for science, but to feed advice into the negotiations on issues such as environmental regulations, product standards and health and safety. "Whether to appoint a CSA is the kind of thought process they should be going through," says Hillman. "It doesn't mean they are there yet, though." ■



Daniel Himmelstein, pictured at his previous research post at the University of California, San Francisco.

INTELLECTUAL PROPERTY

Legal maze threatens to slow data science

Researcher who spent months chasing permission to republish online data sets urges others to read up on the law.

BY SIMON OXENHAM

Knowledge from millions of biological studies encoded into one network — that is Daniel Himmelstein's alluring description of Hetionet, a free online resource that melds data from 28 public sources on links between drugs, genes and diseases. But for a product built on public information, obtaining legal permissions has been surprisingly tough.

When Himmelstein, a data scientist at the University of Pennsylvania in Philadelphia, contacted researchers for permission to reproduce their work openly, several said they were surprised that he had to ask. "It never really crossed my mind that licensing is an issue here," says Jörg Menche, a bioinformatician at the Research Center for Molecular Medicine of the Austrian Academy of Sciences in Vienna.

Menche rapidly gave consent — but not everyone was so helpful. One research group never replied to Himmelstein, and three replied without clearing up the legal confusion. Ultimately, Himmelstein published the final version of Hetionet in July — minus one data set whose licence forbids redistribution, but including the three that he still lacks clear

permission to republish. The tangle shows that many researchers don't understand that simply posting a data set publicly doesn't mean others can legally republish it, says Himmelstein.

The confusion has the power to slow down science, he says, because researchers will be discouraged from combining data sets into more useful resources. It will also become increasingly problematic as scientists publish more information online. "Science is becoming more and more dependent on reusing data," Himmelstein says.

DATA-SET LAWS

Because a piece of data — a fact — cannot be copyrighted, many scientists think that a publicly posted data set that does not place explicit terms and conditions on access can simply be republished without legal problems. But that's not necessarily correct, says Estelle Derclaye, a specialist in intellectual-property law at the University of Nottingham, UK.

The European Union assigns specific database rights, independent of copyright, that aim to protect the investment made in compiling a database. Legally speaking, these rights prevent researchers such as Himmelstein from

STEVE BABULIAK

republishing data sets created by scientists in EU states without their consent.

Other countries have different layers of legal protection. But even in jurisdictions such as the United States, where no separate rights exist to govern databases, there is still room for confusion. Although facts don't qualify for copyright, the way they are compiled arguably might — if the act of making that compilation requires sufficiently creative expression. “The default legal position on how data may be used in any given context is hard to untangle,” according to a guide on licensing data issued by the Digital Curation Centre in Edinburgh, UK.

Advocates of data-sharing accordingly recommend that researchers who are creating public databases add clear licences explaining how they intend their data to be reused and redistributed, and whether they waive any database rights.

LACK OF CONFIDENCE

In Himmelstein's case, some of the data sets that he wanted to use had clear licences — and some of these prevented unrestricted redistribution, but others did not. The most frustrating part of his project, he says, was the feeling that good data were going to waste because their creators could not clarify whether he could republish them.

Andrew Charlesworth, an intellectual-property expert at the University of Bristol, UK, says that this may be because few researchers were confident enough of the law to give Himmelstein clear guidance. “What you tend to find is that if nobody has a remit to answer those kinds of questions, they are not in a hurry to take it on,” he says.

Even without clear permissions, Himmelstein is unlikely to face legal penalties for publishing

“These are largely untested waters, and most academics aren't in the position to risk setting off a legal battle.”

Hetionet, says Jonathan Band, an intellectual-property lawyer with the law firm Policy Bandwidth in Washington DC — unless, that is, he mistakenly breached terms and conditions placed on the data sets. Academics who put their data sets publicly online usually intend their work to be available for others to republish freely; and no one has ever got into trouble for doing Himmelstein's kind of project, Band adds.

But Himmelstein is not convinced that he is legally in the clear — and feels that such uncertainty may deter other scientists from reproducing academic data. If a researcher launches a commercial product that is based on public data sets, he adds, the stakes of not

having clear licensing are likely to rise. “I think these are largely untested waters, and most academics aren't in the position to risk setting off a legal battle that will help clarify these issues,” he says. ■

CORRECTIONS

The News Feature ‘Physics on two wheels’ (*Nature* **535**, 338–341; 2016) contained several biographical inaccuracies. Michael Papadopoulos moved his family to the United States more than a decade before taking a job at Oregon, not in 1967. Jim Papadopoulos spent a whole academic year at Oregon before starting at MIT. He did not write to bike companies asking for work until the 1990s. His time at the US Geological Survey was part of an internship, not a full-time job. The e-mail list he moderated was also founded by him, and is called Hardcore Bicycle Science. He has actually published three first-author papers, but just one related to bicycle science. He was also not given a chance to respond to a comment about his ability to finish things.

The News Feature ‘The beer geeks’ (*Nature* **535**, 484–486; 2016) misattributed the quotes in the last paragraph. They came from Kevin Verstrepen, not Stijn Mertens.

SCHOLARLY OLYMPICS

HOW THE GAMES HAVE SHAPED RESEARCH

BY RAMIN SKIBBA, DANIEL CRESSEY
AND RICHARD VAN NOORDEN
DESIGN BY JASIEK KRZYSZTOFIAK

Whether it's drug scandals, pollution problems or sheer curiosity at the incredible capabilities of the athletes, the Olympic Games have long fascinated researchers as well as the general public.

In recent decades, research has increased on the selection of Olympic sites, environmental issues and the Games' ability to encourage people to participate in sport, says sports-medicine specialist Lars Engebretsen, who heads science and research for the International Olympic Committee.

The Olympics don't typically inspire researchers to start new fields — instead, they tend to feed into ongoing studies, says Vanessa Heggie, a historian of science and sports medicine at the University of Birmingham, UK.

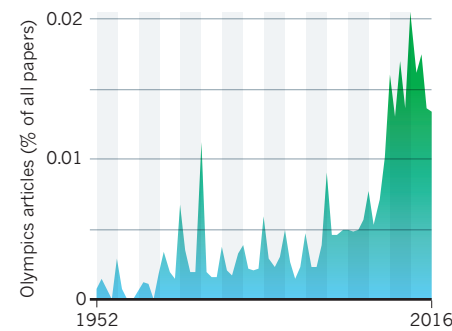
As the 2016 summer Games kick off in Rio de Janeiro, *Nature* uses bibliometrics to provide insight into the who, where, what, how and why of Olympic science.

METHODOLOGY

Using the Scopus database, *Nature* searched for articles that have "Olympics" or "Olympic Games" in the title or "Olympic Games" in the abstract.

SURGE IN SCIENCE

The proportion of research papers that are about the Games has risen rapidly. Over the past few decades, the Olympics has also expanded the number of events, drawn more participants and become vastly more expensive.

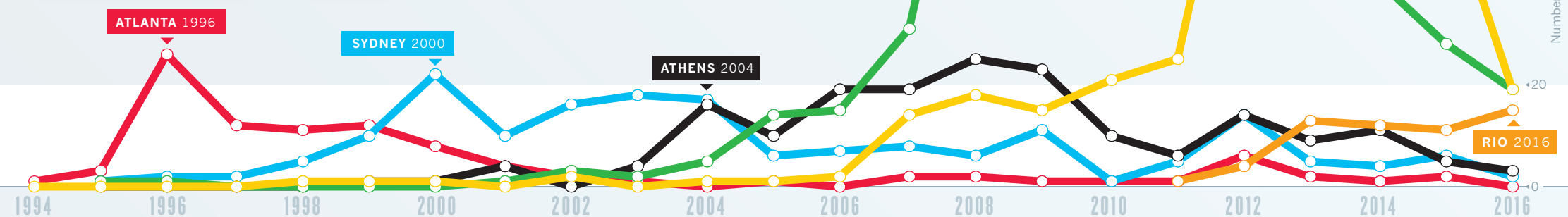


PAPERS PER GAMES

Beijing 2008 inspired the most papers, followed by London 2012.

Beijing had imposed special restrictions on air pollutants, providing a rare opportunity for researchers to do relatively controlled experiments, says David Rich, an environmental epidemiologist at the University of Rochester in New York.

The London 2012 Olympics inspired topics ranging from urban development and sprawl to security and surveillance.



THE DISCIPLINES COMPETE

The social sciences have generated the most Olympics papers — with medicine and engineering winning silver and bronze, respectively.

The Olympics are an "urban change-maker", says sociologist Jacqueline Kennelly at Carleton University in Ottawa, Canada. They have led to expensive infrastructure projects and placed huge demands on public transport. And those that have contended with world wars, protests, boycotts and terrorist attacks have generated substantial literature.

Social scientists have also used the Games to study diverse topics such as the relationships between athletes and coaches (R. A. Philippe and R. Seller *Psychol. Sport Exerc.* **7**, 159–171; 2006) and how much the medal count influences national pride (I. van Hilvoorde *et al. Int. Rev. Sociol. Sport* **45**, 87–102; 2010).

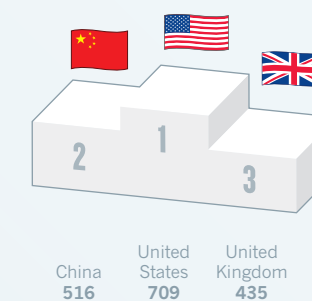
GREECE TAKES GOLD

The countries that have published the most Olympics research are the usual science powerhouses. But divide the number of Olympics papers by the total number of papers published by that country — and different nations take the lead, with Greece at the front of the pack.

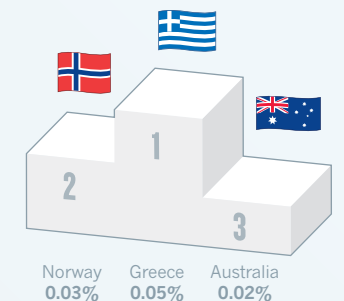
The Olympic Games date back to the eighth century BC, and Greek scientists are naturally proud of their heritage, says Minas Samatas, a political sociologist at the University of Crete in Rethymno, who studied the 2004 Olympics in Athens.

Norway boasts the second highest fraction of Olympics papers, and has won the most medals in the winter Games. Most of its 61 papers are about the winter Games or winter sports, especially skiing.

NUMBER OF OLYMPICS PAPERS



OLYMPICS PAPERS AS PROPORTION OF ALL ARTICLES FOR NATION



CITATIONS IN THE CITY

The paper that has generated the most citations focuses on the Atlanta 1996 Games, and is followed closely by one about Beijing 2008. Both articles explore how policies such as increased provision of public transportation can improve air quality.

The fifth most highly cited paper analysed levels of enthusiasm about the 2000 Olympics among different resident groups in the host city, Sydney. It is the most highly cited Olympics paper in the social sciences.



▶ *FASTER HIGHER STRONGER*

The Cybathlon is a cyborg Olympics that will help disabled people to navigate the most difficult course of all: the everyday world.

BY SARA REARDON

Vance Bergeron was once an amateur cyclist who rode 7,000 kilometres per year — much of it on steep climbs in the Alps. But in February 2013, as the 50-year-old chemical engineer was biking to work at the École Normale Supérieure in Lyons, France, he was hit by a car. The impact sent him flying through the air and onto his head, breaking his neck. When he woke, he learnt that he would never again move his legs on his own, and would have only limited use of his arms.

Confined to bed for months while his body did what healing it could, Bergeron began to look for a way back to cycling. He started to study neuroscience, with an emphasis on research into robotic prostheses that could turn people like him into ‘cyborgs’: combinations of human and machine. He learnt that some of these prostheses used a technique known as functional electrical stimulation (FES) to deliver electrical signals to atrophied limbs or the stumps of missing ones, causing the muscles to contract and restoring some function.

As soon as Bergeron had recovered enough to use a wheelchair, he took that idea back to the lab, where he switched his research focus to neuroscience. Using himself as a guinea pig, he and his team worked out how to stimulate

the nerves in his legs so that his muscles would flex and pedal a bike. “I have become my own research project and it’s a win-win,” he says.

Even with regular exercise sessions to build muscle, Bergeron’s artificially stimulated legs have produced at most 20 watts of power, barely one-tenth of the 150–200 watts produced by an average cyclist. But he and his team are building the FES controller and electrodes into a carbon-fibre recumbent tricycle that he hopes will help him to do better — and perhaps win a medal on 8 October, when he takes his machine to Zurich, Switzerland, to race against other FES cyclists in the Cybathlon: the first cyborg Olympics.

MACHINE LEARNING

Around the world, nearly 80 research groups in 25 countries are honing their technologies for the €5-million (US\$5.5-million) event. They range from small, ad hoc teams to the world’s largest manufacturers of advanced prostheses, and comprise about 300 scientists, engineers, support staff and competitors: disabled people who will each compete in one of six events that will challenge their ability to tackle the chores of daily life. A race for prosthetic-arm users will be won by the first cyborg to complete tasks including preparing a meal and hanging clothes on a line. A powered-wheelchair race will test

how well participants can navigate everyday obstacles such as bumps and stairs.

The venue — Zurich’s 7,600-spectator ice-hockey stadium — should combine with the presence of television cameras and team jerseys to give the Cybathlon a sporting vibe similar to that of the Paralympics, in which disabled athletes compete using wheelchairs, running blades and other assistive technologies. The difference is that the Paralympics celebrates exclusively human performance: athletes must use commercially available devices that run on muscle power alone. But the Cybathlon honours technology and innovation. Its champions will use powered prostheses, often straight out of the lab, and are called pilots rather than athletes. The hope is that devices trialled in the games will accelerate technology development and eventually be used by people around the world.

The everyday tasks being tested in the Cybathlon are much more difficult than they seem, says Robert Riener, a biomedical engineer at the Swiss Federal Institute of Technology in Zurich and creator of the Cybathlon. “I think that people are spoiled by the Internet and Hollywood movies,” he says. “We want to show people there are still challenges.”

Riener traces the origins of the Cybathlon back to news accounts of a charity event: in



Pilot Matt Standridge will compete in the Cybathlon using an exoskeleton designed to help people with paraplegia to walk.

November 2012, a man called Zac Vawter, who had lost a leg in a motorcycle accident, used an experimental, motorized prosthetic leg to climb the 103 storeys of Willis Tower in Chicago, Illinois, in just 45 minutes.

The feat impressed the media — and Riener. But it also frustrated him: Vawter's device, along with similarly impressive prostheses from Riener's lab and others around the world, were not reaching people who needed them. "We're doing great work but not selling it well," he thought. So why not take inspiration from the Willis Tower stunt, and draw attention to the technology through a competition open to everyone in the prosthetics-research community?

Riener's 30-person lab team was excited about designing such an event. And before long, word had spread to colleagues around the world.

At first, Riener had considered hosting showy events such as climbing a mountain on prosthetic limbs. But he changed his mind in 2013, after talking to an acquaintance who had lost an arm to cancer and wore a prosthesis. The device ended in a hook that was moved by cables when the man flexed certain muscles in his stump. It worked well enough for large movements, but was hopeless for fine control. Once, the man told Riener, he had been

buying cinema tickets, and could feel the people queuing behind him staring and growing impatient as he struggled to draw out his wallet and grasp the pieces of paper.

These mundane challenges, Riener realized, were greater and more meaningful than the need to design, say, a spring-like leg that simply helps someone to run fast. So he decided that most of the Cybathlon competitions would be distinctly non-Olympian.

BRAIN POWER

Easily the strangest will be the brain-computer interface (BCI) race, which will feature 15 pilots sitting still for 4 minutes while large screens in the arena show what is going on in their heads. Each will attempt to guide an on-screen character through an obstacle course using specific patterns of brain activity, translated by an electrode cap into three commands: accelerate, jump over spikes or roll under laser rays.

In principle, the patterns can be anything. At the University of Essex in Colchester, UK, for instance, a team of current and former students led by postdoc Ana Matran-Fernandez has designed an algorithm that associates the three motions with a pilot thinking of his or her hand or foot, or working through a maths equation.

The electrical signals are weak, and each

individual is different, so it can be difficult to distinguish between the commands — especially when a pilot is distracted, for example by cheering and adrenaline in the competition. Constantly thinking about tasks is mentally exhausting, says neuroscientist José del R. Millán of the Swiss Federal Institute of Technology in Lausanne, whose team is working on ways to predict thought patterns to make the association more natural and let the pilot relax.

BCIs will probably never be used for real jumping and running, because detecting electrical activity in muscles is much easier. But if such devices could be made cheap and accurate enough, they could help disabled people to guide wheelchairs, cursors or even Skype-enabled robots that would let them participate virtually in an event. "The fact that you can develop this in the lab and bring it out and see it works means there's a future," says Matran-Fernandez.

Other Cybathlon events will highlight the great strides being made with more-conventional devices. In the prosthetic-leg race, competitors must get past obstacles such as stairs, randomly placed stones, tilted pavements and doors — not to mention sitting down in a chair and standing up again. Several participants will be using state-of-the-art smart knees and ankles that can detect force and acceleration as

they walk, and correct their motion if they start to fall.

But even the most advanced engineering pales beside what the intact body does naturally. When a person picks up a pen with a flesh and blood arm, their brain and peripheral nervous system coordinate how far to reach, how to bend each joint in each finger into a precise shape, and how hard to grasp — all without conscious effort. Standard movable prostheses, such as the type with the hook and cables, require the user to do all of this consciously. This takes a great effort, which is one reason many amputees choose not to wear them.

To get around that, researchers have to create computer algorithms that decode signals from muscles and nerves and predict what a wearer is trying to do. In Burnaby, Canada, a Cybathlon team called MASS Impact is working with pilot Danny Letain, a former Canadian Paralympic skier who lost his left arm in a 1980 railway accident. The team has built an arm with a panel of flat buttons that sits on Letain's arm stump.

Using his memory of a hand, Letain imagines making one of 11 gestures, such as pointing. The muscles in his stump then compress the buttons and tell his artificial hand to do what he intends. Letain was pleased to find that the brain circuitry that once controlled his fingers is still in working order, long after he stopped feeling any 'phantom pain' in his lost arm. "I'm using something I haven't used in 35 years," he says.

Some arms are even more advanced. A team led by biomedical engineer Max Ortiz Catalan at Chalmers University of Technology in Gothenburg, Sweden, has developed a two-way prosthetic hand that can feel as well as move (M. Ortiz-Catalan *et al. Sci. Transl. Med.* 6, 257re6; 2014). The arm is permanently implanted in the wearer's bone, and uses up to nine electrodes to convey motor commands from the remaining muscles to the prosthesis, and to send signals from sensors in the fingers back to the arm's sensory nerves. Cybathlon pilot Magnus Niska is the only person in the world who wears such a prosthesis outside the lab. Ortiz Catalan hopes that the ability to feel objects will give Niska a competitive advantage.

A team led by Ronald Triolo, a biomedical engineer at Case Western Reserve University in Cleveland, Ohio, has a similar strategy for the FES cycling event, in which contestants with spinal-cord injuries will pedal for 750 metres around a circular track. Many of the competitors, including Bergeron, use electrodes on the skin to stimulate the leg muscles. But the Cleveland system — originally designed to allow people with lower-limb paralysis to walk with the help of crutches — features electrodes surgically implanted in the leg muscles. Using an external device, the wearer chooses a



Functional electrical stimulation helps people with spinal injuries race bikes.

menu option, such as 'sit'. An implanted pulse generator activates the electrodes that cause the muscles to contract in the correct order.

After Triolo heard about the Cybathlon, he realized that he could add cycling to his volunteers' exercise regimes. His team has equipped a recumbent tricycle with sensors that detect the angle of the cyclist's leg as he or she pedals, and automatically change the stimulation patterns so that one leg pushes while the other pulls.

Triolo says that all 27 of the people implanted with his electrodes want to try cycling. After putting them through qualifying trials, he is down to a few finalists for the Cybathlon. "We want to go to Switzerland and win this thing," he says. "Then I'd like to use that as a springboard to build an exercise programme here."

EYES ON THE PRIZE

This competitiveness is a far cry from Triolo's initial reaction to the Cybathlon, which was that the competition was a foolish idea. "We should find a way to collaborate internationally on these problems rather than compete," he recalls saying — something that he and Riener say they still hear from some in the prosthetics field.

Triolo eventually came around: he decided that the Cybathlon would at least be a good learning experience. Riener himself hopes that bringing the competition into the open will spur creativity better than the conventional academic process, which is hampered by researchers' concerns about their intellectual property and competitiveness for grants.

Karim Lakhani, an economist who studies innovation at the Harvard Business School in Boston, Massachusetts, notes that competitions also force researchers to finish their work quickly and eliminate doubts about feasibility that prevent them from starting in the first place. He points to the self-driving car, which languished in development for decades until 2005, when the US Defense Advanced Research Projects Agency held a race with a \$2-million prize. The contest eventually drew interest from Google, which is now testing such cars on

the roads. "This contest will serve the same way," says Lakhani. The Cybathlon will not award monetary prizes, just medals. But his research suggests that the recognition enjoyed by the winners could be just as motivating (K. J. Boudreau *et al. RAND J. Econ.* 47, 140–165; 2016).

Perhaps the greatest advantage of prizes is that they give unknown contestants an opportunity to compete alongside big, well-known players, says Lakhani. The Cybathlon has drawn plenty of both. Otto Bock HealthCare, a multibillion-euro company based in Duderstadt, Germany, and the world's largest manufacturer of prosthetic limbs, has entered three Cybathlon events.

One is the powered-exoskeleton race, in which contestants with spinal injuries will use an external support system to navigate obstacles similar to those in the powered-leg race. Otto Bock's pilot, Lucia Kurs, lost the use of her legs to spinal tumours. Now in her 60s, she can walk 12 kilometres using the firm's commercial leg brace, which has sensors, electronics and motors to guide the knees and ankles through a normal leg swing.

"We're showing off, and checking out other manufacturers" at the Cybathlon, says Christof Küspert, a product manager at Otto Bock. But he says the company is also interested in learning about innovative prototypes from dark-horse developers at universities.

One smaller developer is Jesús Tamez-Duque, managing director of the start-up INDI Engineering and Design in Monterrey, Mexico, who is entering a prototype for an exoskeleton much cheaper than Otto Bock's US\$75,000 model. The device's joints are moved by wind-screen-wiper motors, and much of the body is 3D printed. INDI's competitor uses a joystick attached to his crutches to choose from several programmed movements, such as climbing stairs or sitting.

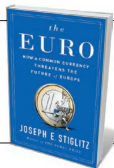
Tamez-Duque hopes that the Cybathlon will attract collaborators and prove that Mexico can be a player in the field. "The way we see it, the Cybathlon is a competition that concentrates the top-notch robotics labs in the world," he says. "We're still working on getting that representation so that other people believe we can actually add value to them."

The Cybathlon will be back in 2020, as a seven-day event in Tokyo, coinciding with the Olympics. It will have new events for competitors with visual impairments, and will conduct some races outside the stadium. But for competitors in the first Cybathlon, being at the cutting edge is already a thrill. "This is *Iron Man*, this is *Avatar*," says Bergeron. "It's a combination of BCI and exoskeletons all over the place." ■

Sara Reardon is a staff reporter for Nature in Washington DC.

COMMENT

ECONOMICS Joseph Stiglitz's guide to the failures of eurozone policymakers **p.26**



DEVELOPMENT Parenting tips from the lab of a 'Berkley bubble' **p.27**

CHINA A bold and epic portrait of a civilization shaped by water **p.28**

CAPACITY-BUILDING Mentoring scheme supports female scientists in Cameroon **p.30**

JIE ZHAO/CORBIS/GETTY



An electronic-waste recycling factory in Hubei, China.

Take responsibility for electronic-waste disposal

International cooperation is needed to stop developed nations simply offloading defunct electronics on developing countries, argue **Zhaohua Wang, Bin Zhang and Dabo Guan**.

The world is producing ever more electrical and electronic waste. The quantity of dumped computers, telephones, televisions and appliances doubled between 2009 and 2014, to 42 million tonnes per year globally^{1,2}. Developed countries, especially in North America and Europe, produce

the most e-waste (see 'Unfair flow'). The United States generates the largest amount, and China the second most³.

Much of this waste ends up in the developing world, where regulation is lax. China processed about 70% of the world's e-waste in 2012⁴; the rest goes to India and other countries in eastern Asia and Africa,

including Nigeria⁵. Non-toxic components — such as iron, steel, copper and gold — are valuable, so are more frequently recycled than toxic ones⁴. Disposal plants release toxic materials, volatile organic chemicals and heavy metals, which can harm the environment and human health.

Lead levels sampled in the blood ►

► of children in the e-waste-processing town of Guiyu, China, were on average three times the safe limit recommended by the US Centers for Disease Control and Prevention⁶. In California, peregrine falcons have been threatened — polybrominated diphenyl ethers, which are widely used as flame-retardants in electronics, have been discovered in their eggs.

A global approach to managing the volume and flow of e-waste is urgently needed. This requires: an international protocol on e-waste; funding for technology transfer; firmer national legislation on imports and exports; and greater awareness of the problem among consumers. Researchers and regulators should build a global e-waste flow system that covers the whole life cycle of electrical goods, including production, usage, disposal, recovery and remanufacturing.

Beyond better recycling, the ultimate aim should be a circular economy of cleaner production and less wasteful consumption, including the embrace of a sharing economy and cloud-based technologies with smaller material footprints. As the world's largest producer of electronic goods and recipient of the most e-waste, China should take the lead.

BAD RUBBISH

Most developed countries have strict regulations governing the disposal of electronic and electrical waste. European countries, the United States and others have official 'take-back' systems, which recover and dispose of e-waste in an environmentally friendly way. In 2014, these processed 6.5 million tonnes generated by 4 billion people, recycling valuable materials back into the supply chain. The European Union has two comprehensive directives: the Restriction of Hazardous Substances and Waste Electrical and Electronic Equipment. Yet the EU and the United States and Canada dispose domestically of only 40% and 12%, respectively, of the e-waste they generate³.

These rich nations with strict legislation send most of their e-waste to developing countries. India and China's e-waste legislation is inefficient and irregularly enforced. China's system is poorly coordinated; it involves more than ten departments publishing regulations, imposing disposal fees, providing subsidies and monitoring pollution and illegal imports with little crosstalk. Many poor nations, especially in Africa, have few or no laws on e-waste.

Around half the components in any personal computer contain mercury, arsenic and chromium — all are toxic. The movement of this waste in and out of countries is not being tracked. The Basel Convention



A worker dismantles a motherboard by hand in Sangrampur, India.

of the United Nations, which concerns the movement of hazardous waste across borders, is meant to prevent developed countries from illegally dumping hazardous waste in developing countries. But only 87 parties — and not the United States — have ratified it. Few developing countries control imports of toxic e-waste: for example, India's law fails to ban it. This resulted in 50,000 tonnes of such waste from developed countries being dumped in India in 2012⁷. The shady trading of trash as 'used electronics' bypasses such laws entirely.

In many developed countries, such as those in the EU, manufacturers are required to take responsibility for the disposal of their electrical and electronic products. However, three-quarters of products sold in Europe are made in developing countries such as China and India. So such measures only worsen the situation in poor nations.

UNCERTIFIED DISPOSAL

A few developing countries, including China, have made producers responsible for some disposal. Since January 2011, Chinese producers have had to pay disposal fees for five categories of home appliance

(televisions, air conditioners, refrigerators, washing machines and computers). The list grew to 14 in March this year. But the scheme pays only for e-waste processing, not collection — so e-waste is just not collected.

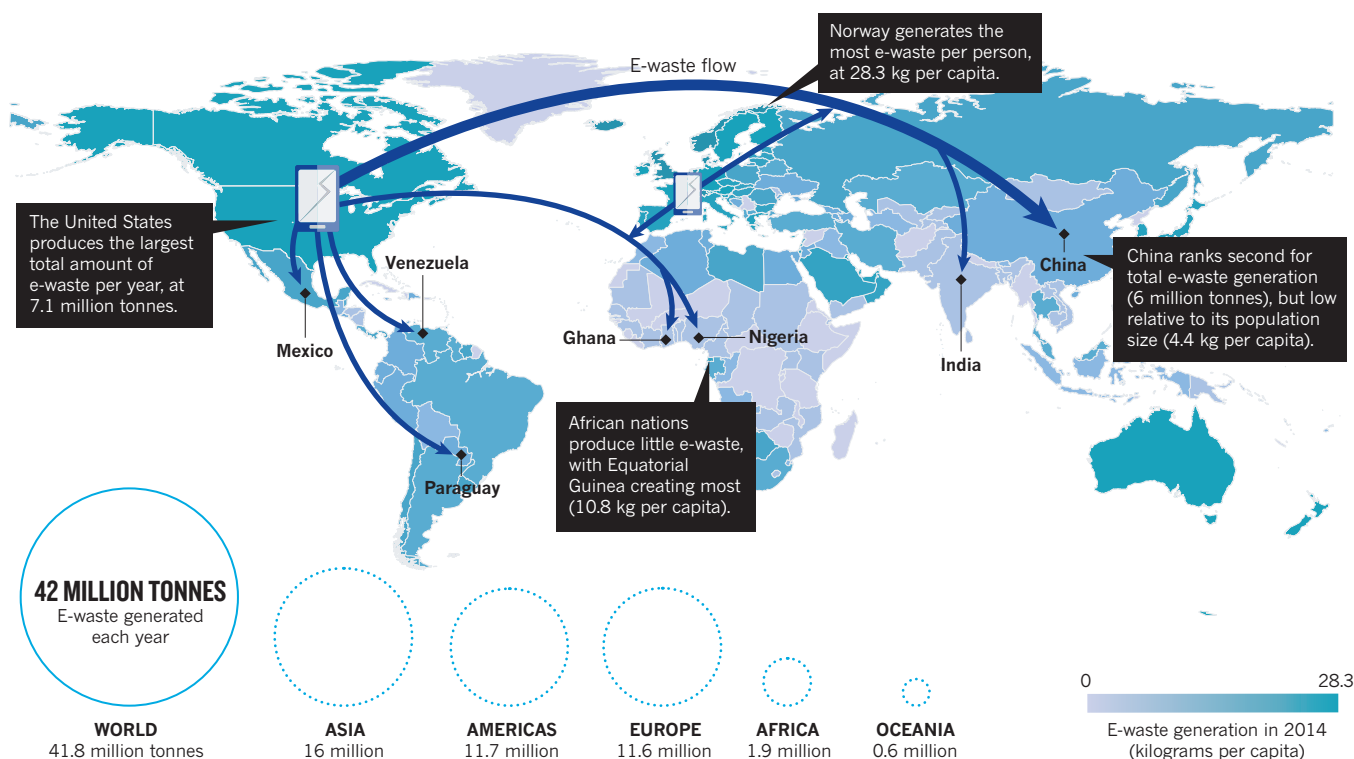
China has 106 enterprises certified by the government as capable of dismantling 100 million defunct home appliances per year. Together, these companies process only 40 million items. The rest is recovered by unskilled peddlers going door-to-door⁸. There are 300,000 such people in Beijing alone. They sell to uncertified disposal plants⁹; these pay higher prices than the certified ones, which have larger overheads. Elsewhere — in Guiyu, for example, or in Agbogbloshie in Ghana — people sift e-waste from household rubbish and sell it for uncertified disposal. In China, this 'grey' market is estimated to be worth US\$15 billion. Large amounts of government subsidies intended for such disposal lie idle. In 2013, only 630 million yuan (US\$94 million) of the 2.81 billion yuan available was spent.

By contrast, EU-based manufacturers are encouraged, through legislation, to support the dismantling of their products and the recovery of materials through environmentally friendly product design¹⁰. In Japan, producers must collect used devices; consumers are required to deliver their e-waste to the manufacturer or to certified collection sites, and pay a compulsory recycling

SANJIT DAS/BLOOMBERG/GETTY

UNFAIR FLOW

Most electronic waste from developed countries ends up in poor nations that lack regulation. China processed around 70% of the world's e-waste in 2012; the rest goes to India and other countries in eastern Asia and Africa, including Nigeria.



SOURCE: REF. 3

fee to the waste-disposal company.

Globally, only 6.5 million tonnes of e-waste (about 15% of the total) were formally reported as disposed of through national take-back systems in 2014. In China, the proportion is 24–30%³. The rest is sent either to landfill or to the black market.

FOUR STEPS

The following four steps need to be taken to make e-waste management sustainable.

First, a formal global protocol on e-waste trading needs to be built under the Basel Convention, and the United States must be encouraged to participate. The convention currently covers only the trading of toxic waste; it should be extended to encompass e-waste and second-hand electronic products. Strict criteria must be agreed globally to distinguish products by durability, usability and safety.

Second, domestic regulations need strengthening and enforcing: those operating illegally should be fined or prosecuted. Developed countries must crack down on defunct products being traded as used ones. Developing countries must ban imports of toxic e-waste. Customs duties on e-waste should be increased.

Third, the United Nations' Solving the E-waste Problem Initiative must take on many more roles. It should launch a global industry association to certify

processing firms that meet agreed legal, technical and environmental criteria. It should encourage the transfer of processing and recycling technology from developed to developing nations. It should create a global e-waste disposal fund to which exporting countries and manufacturers would contribute for each product they sell.

Any country responsible for disposal should receive a fee that is 2–5% of the original production cost, and ensure that an appropriate and verifiable disposal procedure is implemented. Certified firms would get subsidies from the fund according to how much they process. The same industry body should launch a global monitoring system to track flows around the world over the whole life cycle of e-products. Components such as circuit boards and the compressors in refrigerators and air-conditioners could be labelled with radio-frequency identification tags.

Fourth, consumers' responsibility for e-waste needs to be enshrined in regulations, taking lessons from Japan. Separate e-waste bins should be provided, with penalties for those who do not use them. Deposit mechanisms could be used when purchasing electrical goods, and people can get the money back when they send their waste to certified collectors.

It is time for consumers, researchers, manufacturers, nations and international

regulators to direct some of the passion and creativity they have for new gadgets towards responsibly dealing with old ones. ■

Zhaohua Wang is professor and **Bin Zhang** is a researcher in the School of Management and Economics, and the Center for Energy and Environmental Policy Research, Beijing Institute of Technology, China. **Dabo Guan** is professor in climate-change economics, School of International Development, University of East Anglia, Norwich, UK. e-mail: dabo.guan@uea.ac.uk

1. Robinson, B. H. *Sci. Tot. Environ.* **408**, 183–191 (2009).
2. Baldé, C. P. et al. *E-waste Statistics: Guidelines on Classifications, Reporting and Indicators* (United Nations University, IAS-SCYCLE, 2015).
3. Baldé, C., Wang, F., Kuehr, R. & Huisman, J. *The Global E-waste Monitor-2014* (United Nations University, IAS-SCYCLE, 2015).
4. Zhang, K., Schnoor, J. L. & Zeng, E. Y. *Environ. Sci. Tech.* **46**, 10861–10867 (2012).
5. Breivik, K., Armitage, J. M., Wania, F. & Jones, K. C. *Environ. Sci. Tech.* **48**, 8735–8743 (2014).
6. Ogunseitan, O. A., Schoenung, J. M., Saphores, J.-D. M. & Shapiro, A. A. *Science* **326**, 670–671, (2009).
7. Agoramoorthy, G. & Chakraborty, C. *Nature* **485**, 309 (2012).
8. Liu, X., Tanaka, M. & Matsui, Y. *Waste Mgmt Res.* **24**, 434–445 (2006).
9. Wang, Z., Zhang, B., Yin, J. & Zhang, X. *J. Clean. Product.* **19**, 977–984 (2011).
10. European Commission. 'Directive 2002/96/EC of the European parliament and of the council on waste electrical and electronic equipment.' *Off. J. Eur. Commis.* (2003); available at <http://go.nature.com/2at98Jq>



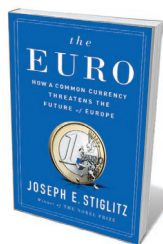
Austerity measures in Greece — part of the eurozone — have sparked unease among pensioners.

ECONOMICS

Singular currency

Jonathan Portes parses Joseph Stiglitz's analysis of the euro in the context of the global financial crisis.

Who killed the euro? With the spotlight now on Britain's relationship with the currency's crucible, the European Union, that question has an edge. Were the culprits the visionary politicians who ignored structural differences between European economies when they conceived the euro in the 1980s — most notably Jacques Delors, president of the European Commission at the time? Was it the neoliberal international economic and financial establishment, in Washington DC, New York and London as well as Frankfurt and Brussels, that thought markets (even financial ones) were rational and self-correcting? Was it Jean-Claude Trichet, governor of the European Central Bank during the 2008 financial crisis, who made the people of Ireland pay for the corruption of their bankers by forcing their government to stand behind their broken banks? Or perhaps the putative 'Swabian housewife' often invoked by German Chancellor Angela Merkel, who balances her budget and cannot see why the



The Euro: How a Common Currency Threatens the Future of Europe
JOSEPH E. STIGLITZ
W. W. Norton: 2016.

of a century. Stiglitz speculates boldly and cogently about the zone's future, but *The Euro* is on the whole too scattershot either to diagnose the illness or to prescribe a viable cure.

Stiglitz's thesis — or rather, loose assemblage of theses — explains too much. The gross domestic product of the eurozone is now barely higher than it was immediately before the crisis, and the euro is indeed

same logic does not apply to governments?

In the first chapter of economist Joseph Stiglitz's *The Euro*, we learn that they all did it. And that it is all the fault of contemporary capitalism. Most of the book is devoted to this topic, and it serves as a useful guide to the many failures of eurozone policymakers in the past quarter

central to explaining why. There is, however, a lively debate over the inevitability of this failure. Some argue that the euro was always doomed because of divergences in income and economic structures between eurozone countries, and the lack of a US-style political union to make significant fiscal transfers possible. Sooner or later, a big shock — a rapid rise in unemployment, a sharp fall in capital inflows — would hit one country or more, and without being able to devalue their currency, nations would be unable to adjust without excessive economic damage. It was bad luck that when the shock did come (in 2008) it was as big as it was, especially for countries such as Spain and Ireland. It was also both predictable and predicted.

Others, such as economics journalist Martin Sandbu in his excellent *Europe's Orphan* (Princeton Univ. Press, 2015), argue that the eurozone's dismal economic performance is a result less of the currency's design than of a set of avoidable and entirely unnecessary policy failures. These include premature and excessive austerity measures, especially in southern European countries, and slowness in restructuring debts, in particular the Greek government's. Stiglitz leans towards the view that failure was inevitable. Unlike Sandbu, he doesn't provide much in the way of evidence one way or the other.

Similarly, he fails to explain what he thinks has happened outside the eurozone. It's true, as he shows, that the eurozone's productivity performance since 2008 has been abysmal. But productivity has grown even more slowly in Britain (which held onto its own currency), although employment has held up better. A more nuanced account would distinguish between two sets of factors. That is, those that have been present in most advanced countries to some degree (including excessive austerity and the still-unexplained productivity slowdown), and those specific to the eurozone and its institutions — in particular, the inability to adjust exchange rates, and the lack of a central bank able to act as a credible lender of last resort.

Stiglitz also speculates about the near future. Unsurprisingly, given the shopping list of flaws that he sees in the euro, he thinks that radical reform is necessary for it to survive. That includes fiscal policy that is both more expansionary — with greater scope to increase public investment — and more countercyclical, so that spending can rise rather than fall in a recession. It would also include greater regulation, especially for the financial sector and a new mandate for the European Central Bank that focuses less heavily on price stability and more on growth and employment. Yet Stiglitz fails to distinguish clearly between reforms that are essential to save the euro and those that would be 'nice to have', and that, I would bet, he would recommend equally to non-eurozone countries

such as Britain or the United States.

If such reforms do not happen, Stiglitz recommends either an “amicable divorce” — dissolution of the eurozone — or a move to a more “flexible euro”. This is the most innovative and interesting part of the book. He argues for a new, and much more heavily regulated and controlled, international monetary system. States would have much more direct control of both money creation internally and their current account balances externally. He advocates market-based mechanisms for both. “Credit auctions” would demand that private banks pay for the right to expand the money supply, and “trade chits” would force importers to effectively buy tradeable licences to import. Nevertheless, this would be a radical move towards greater state control of the economy.

This is potentially very exciting. Few would defend the current organization of the international financial system, and radical ideas based in sound economics are exactly what we need. The state control that Stiglitz is advocating would be viewed with scepticism

“Radical ideas based in sound economics are exactly what we need.”

at the International Monetary Fund or the US Treasury, but would no longer be regarded as laughable or heresy. And a proposal from someone of Stiglitz’s eminence has a good claim on our attention. Unfortunately, this part of the book is underdeveloped. For example, it seems unlikely that as good an economist as Stiglitz hasn’t thought about how the spread of shadow banking — borrowing and lending outside the traditional banking system — has made it much more difficult to control credit creation. And he must be aware of the administrative problems that trade in services (particularly tourism) would pose for his chit system. But these issues are not addressed.

Will the eurozone respond to Britain’s vote to leave the EU with a rapid move towards greater integration, or will the tensions identified by Stiglitz pull it apart? Perhaps, as has mostly been the case so far, it will continue to muddle through. But it cannot avoid the questions that Stiglitz poses for ever — even if he is a long way from providing convincing answers. ■

Jonathan Portes is principal research fellow at the UK National Institute of Economic and Social Research in London. He was chief economist for the UK Cabinet Office during the 2008–09 financial crisis. e-mail: j.portes@niesr.ac.uk

CHILD DEVELOPMENT

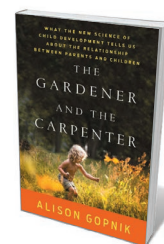
A cognitive case for un-parenting

Josie Glausiusz relishes Alison Gopnik’s study on how child-rearing demands the embrace of messy realities.

An Amazon trawl for “parenting books” last month offered up 186,262 results. Titles included Daniel Siegel and Tina Payne Bryson’s *The Whole-Brain Child: 12 Revolutionary Strategies to Nurture Your Child’s Developing Mind* (Delacorte, 2011), Elaine Glickman’s *Your Kid’s a Brat and It’s All Your Fault* (TarcherPerigee, 2016) and *Have a New Kid by Friday* by Kevin Leman (Revell, 2012). This is less genre than tsunami.

Yet, as Alison Gopnik notes in her deeply researched book *The Gardener and the Carpenter*, the word parenting became common only in the 1970s, rising in popularity as traditional sources of wisdom about child-rearing — large extended families, for example — fell away. Gopnik, a developmental psychologist (or as she describes herself, “a bubble at Berkeley, a grandmother who runs a cognitive science laboratory”), argues that the message of this massive modern industry is misguided.

It assumes that the ‘right’ parenting techniques or expertise will sculpt your child into a successful adult. But using a scheme to shape material into a product is the modus operandi of a carpenter, whose job it is to make the chair steady or the door true. There is very little empirical evidence, Gopnik says, that “small variations” in what parents do (such as whether they sleep-train) “have reliable and predictable long-term effects on who those children become”. Raising and caring



The Gardener and the Carpenter: What the New Science of Child Development Tells Us About the Relationship Between Parents and Children
ALISON GOPNIK
Farrar, Straus & Giroux: 2016.

for children is more like tending a garden: it involves “a lot of exhausted digging and wallowing in manure” to create a safe, nurturing space in which innovation, adaptability and resilience can thrive. Her approach focuses on helping children to find their own way, even if it isn’t one you’d choose for them. The lengthy childhood of our species gives kids ample opportunity to explore, exploit and experiment before they are turned

out into an unpredictable world.

In Gopnik’s not-parenting approach, the rampant disorder of genetic variation (or, to use her technical term, “mess”) becomes a wellspring for creativity, contributing to the wide range of children’s temperaments and abilities. Some children are risk-takers; others are timid; some are highly focused (an advantage in a test-obsessed school system) or natural hunters (“constantly on the alert for even subtle changes in the environment”). Throughout history, she argues, that mix has bred resilience in societies faced with challenges, such as early nomads’ constant need to confront new environments. People with more conservative temperaments, for example, ensure some security for the risk-takers.

Gopnik reveals how the parenting model can affect how children explore. She describes a wide range of experiments showing that children learn less through “conscious and deliberate teaching” than through watching, listening and imitating. Among the K’iche’ Maya people of Guatemala, even very young children with little formal schooling can master difficult and dangerous adult skills — such as using a machete — by watching adults engaging in these tasks in slow and exaggerated fashion. In one of Gopnik’s own experiments using a “blicket detector” (a box that lights up and plays music when a certain combination of blocks is placed on it) four- and five-year-olds worked out that unusual ▶



Children learn well from undirected play.

► combinations rather than individual blocks did the trick — and younger kids were more skilled than older ones at finding unlikely options.

She also cites a number of studies on play, which is so crucial to human development that children engaged in it even in Nazi concentration camps. Research on dolphins, crows and foxes reveals how playing at hunting, digging and fighting develops the skills the animals need as adults. Through play, young rats produce chemicals called cholinergic transmitters, implicated in plasticity in ‘social’ areas of the brain. Rats deprived of play when young can defend, attack or approach others as adults, but

fail to know “when to do what”, she notes. Most human parents, Gopnik writes, “have a vague sense that play is a Good Thing”. But as an aim of parenting, play is paradoxical, she claims, because it is essentially goalless. Elizabeth Bonawitz, a researcher in computational cognitive development, found that when adults instructed children on how to play with a squeaking toy, the children imitated them. When left to their own devices, the children were more likely to try different actions until they had discovered everything the toy could do.

Gopnik can be scathing in her censure of the modern educational system, which increasingly stresses high-stakes testing. That trend, she notes, parallels the rise in diagnoses of attention deficit hyperactivity disorder (ADHD), which in the United States particularly is often treated with drugs that can have serious side effects, including addiction. More palpable, however, is her devotion to the subjects of her research, including her grandchildren Augie and Georgie, her “true muses”, whose antics pepper her text.

Those antics remind me of my own delightfully disorderly, creative five-year-old twins and their in-the-now mischief and affection. As Gopnik concludes: “The most important rewards of being a parent aren’t your children’s grades and trophies — or even their graduations and weddings. They come from the moment-by-moment physical and psychological joy of being with this particular child, and in that child’s moment-by-moment joy in being with you.” ■

Josie Glausiusz writes about science and the environment for magazines including *Nature*, *National Geographic* and *Hakai*.
Twitter: @josiegz

“Gopnik can be scathing in her censure of the modern educational system.”



The Three Gorges Dam on the Yangtze River is one of the world’s largest power stations.

CHINA

A hydrological history

Andrea Janku enjoys a study of the nation-building role of China’s great rivers, the Yellow and the Yangtze.

Nearly 70 years ago, Chinese anthropologist Fei Xiaotong published *From the Soil* (1947). The Chinese people, he wrote, were “inseparable from the soil”, which had produced “a glorious history”, but one that was “limited by what could be taken from the soil”. If that book was the portrait of a rural and inward-looking country, literally stuck in the famous yellow earth — the loess of the North China Plain — science

writer Philip Ball’s history of China, *The Water Kingdom*, is very much the opposite.

It is the portrait of a civilization permeated by water, with patterns of thought influenced by the centrality of water to everyday life and, echoing that, practical affairs shaped by philosophical ideas based on the principle of flow. The result is, Ball writes, “an intimate connection between hydraulic engineering, governance, moral rectitude and metaphysical



STR/AFP/GETTY

speculation that has no parallel anywhere in the world". On this premise he builds a picture of the nation, from its geographical and ideological foundations to the environmental and political predicament in which China (and not only China) now finds itself.

The Water Kingdom's structure is predominantly thematic rather than chronological. So the first chapter, introducing the Great Rivers, the Yangtze and the Yellow, leads from the Great Yu, the mythical ruler who, according to tradition conquered the floods more than 4,000 years ago, to twentieth-century Communist leader Mao Zedong, who repeatedly reasserted his power by swimming in the Yangtze. It interweaves more stories of the Yangtze: seventeenth-century explorer Xu Xiake's search for its source, twelfth-century poet Lu You's descriptions of commercial life along its banks, and more recent Western visitors' accounts of the colossal Three Gorges Dam.

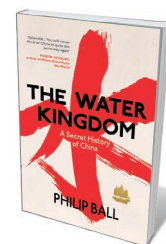
On this epic journey, Ball explores mythological accounts of dragons and floods, along with early philosophical texts such as the teachings of Mencius from the fourth

century BC, to unravel the origins of Chinese political ideology. That is, the idea that he who controls the water controls the people, which links the earliest cultural heroes to modern leaders from Republic of China founder Sun Yat-sen to Mao himself. Ball subsequently covers Zheng He's maritime explorations at the height of Ming-dynasty power in the fifteenth century, and the centrality of the Yellow River–Grand Canal administration to the state bureaucracy in the eighteenth. He notes how the late empire turned into a "hydraulic state", increasingly mired in systemic problems that finally collapsed under the pressure of internal rebellion and the imperialist onslaught.

The centrality of water even plays out in the art of war, to which Ball devotes a chapter. In some of the most dramatic conflicts, rivers were harnessed as weapons. In 204 BC, for instance, an intentional rupturing of the Wei River dams led to the victory of the Han-dynasty forces. And in 1938, the Nationalist government attempted to stop the advancing Japanese army by breaching the Yellow River dykes, with disastrous consequences for the Chinese people — killing hundreds of thousands and making millions homeless. Technological and political parameters changed fundamentally in the twentieth century, yet hydraulic nation-building and the myths that surrounded it assumed an ever more important role. Mao in particular relished the role of the great leader conquering the floods. Ball ends that strand of the narrative with the Three Gorges Dam, first conceived by Sun Yat-sen and finally completed in 2012. He even covers the depiction of water in Chinese art through the ages, exploring its aesthetic, philosophical and political dimensions. The journey ends with a pertinent chapter on China's current environmental crisis. Another hydraulic-engineering project on an unprecedented scale, the South–North Water Transfer Project, is now under way, meant to tackle water scarcity in the north



Mao Zedong by the Yangtze.



The Water Kingdom: A Secret History of China
PHILIP BALL
Bodley Head: 2016.

(J. Barnett *et al. Nature* 527, 295–297; 2015).

Telling the history of Chinese civilization from the perspective of water is rewarding, because it can link the history of ideas and beliefs, technology and warfare, politics and the arts. But as with any general history, it risks essentializing China and making its history seem more uniform than the actual record justifies. The most obvious example is the major shift in the history of the Yellow River in the late tenth century. It is only from then that the river became a constant threat, bursting its dykes and flooding the countryside in ever more devastating cycles, and changing its course repeatedly in dramatic ways after nearly a millennium of relative stability. And so it is also only from then that controlling the river became tantamount to controlling the people, and that state and society became trapped in an increasingly unsustainable hydraulic infrastructure. That complex system of dykes and canals, with the Yellow River and the Grand Canal at its heart, devoured enormous resources — a quandary called "technological lock-in" by historian Mark Elvin. Moreover, Ball's focus on the state means that he fails to mention the role of small-scale irrigation and conservation projects that are common in particular in the southern China, and largely managed and funded by the local gentry.

Still, this is a convincing introduction to Chinese history. Rather than perpetuating stereotypes, it boldly navigates the treacherous and often-avoided terrain long dominated by influential but spurned theories, such as the idea, promoted by sinologist Karl August Wittfogel, of China as a despotic hydraulic society. It also complements and complicates Fei Xiaotong's idea of an earthbound civilization — a metaphor that has had a huge impact in China itself. In 1988, the six-part Chinese television documentary *River Elegy* depicted the country as weak and backward, closed off from the world by the Great Wall and stuck in the mud of the Yellow River, contrasted with a progressive, open, oceanic conceptualization of Western civilization. It has taken a generalist to turn the rich but rather dry literature on the history of water in China into an accessible history. Why it is a secret one, however, remains a mystery. ■

Andrea Janku is senior lecturer in Chinese history at SOAS, University of London. She is currently working on a monograph on the history of famine, *Integrating the Body Politic*.
e-mail: aj7@soas.ac.uk

SONFOTO/UG VIA GETTY

Correspondence

China: change tack to boost basic research

We agree that China must invest more in basic research, but fear that simply casting more seeds on infertile ground will not yield the anticipated fruit of innovation (W. Yang *Nature* **534**, 467–469; 2016). More bottom-up initiatives for early-career researchers are required for long-lasting change.

A priority is to build capacity in critical thinking and self-determination, both of which are cornerstones of creative enquiry. Early-career scientists should be trained and judged on more than just technical competence. Furthermore, strategies are needed to give young researchers in China the same opportunities that leading Western institutions take for granted. Limited access to key information services and an educational emphasis on written knowledge over verbal communication skills do not foster scientific debate.

Structural reform of funding silos and hierarchical power structures in science institutions is essential for cross-sector collaboration — a crucial contributor to scientific progress in Western countries. Chinese funding schemes such as the 10,000 Talents programme could be redistributed across an evened-out power structure, and associate professors allowed to supervise PhDs and lead their own groups.

The applied-research sector receives much more funding than basic science does (see *Nature* **534**, 452–453; 2016), so it could help by promoting the possible benefits of its work for basic research in reports and funding applications. **Raphael K. Didham, Chao-Dong Zhu** *Institute of Zoology, Chinese Academy of Sciences, Beijing, China. zhucd@ioz.ac.cn*

China: standardize R&D costing

Figures for China's basic-research spending should not be taken at face value (see W. Yang *Nature*

534, 467–469; 2016). Before comparing gross expenditure on research and development (R&D) with that in developed countries, China's official statistics first need bringing into line with international standards for collecting and reporting R&D costs (see go.nature.com/2ab54rh).

Unlike most other countries, China's government has no official system for assessing R&D expenditure. These costs are instead embedded in the overall costs for science and technology (see Y. Sun and C. Cao *Science* **345**, 1006–1008; 2014). And China's R&D statistics exclude salaries for university faculty members and postdocs — a significant component of R&D expenditure in other countries. Capital spending on big facilities and their operation budgets, known as fixed R&D costs in many Western countries, are not counted. For instance, only about US\$18 million of the \$200 million spent on the first construction phase of the Shanghai Synchrotron Radiation Facility was designated as an R&D cost (see go.nature.com/2alwvbc).

Based on such considerations, China probably spent about double the official figure of 4.7% of its total R&D budget on basic research in 2013, as quoted by Yang. This is still significantly less than Japan or the United States.

Yutao Sun *Dalian University of Technology, China.*

Cong Cao *The University of Nottingham Ningbo, China. sunyutao82@dlut.edu.cn*

Mentoring female scientists in Africa

As founding members of the Higher Institute for Growth in Health Research (HIGHER) for Women in Cameroon, our mission has been to help young women to enter and sustain careers in biomedical science through a mentoring programme

(www.higherwomenscam.org).

Mentored female researchers spend more time on research and have more publications and greater career satisfaction than do their unmentored peers (W. Levinson *et al. West J. Med.* **154**, 423–426; 1991). In disadvantaged settings such as Cameroon, however, potential female mentors are in short supply.

Led by one of us (R. G. F. L.) and funded by the Special Programme for Research and Training in Tropical Diseases and Canada's International Development Research Centre, the HIGHER Women consortium has recruited more than 100 members in the past year. Its 20 or so mentors hold leading positions in Cameroon in academic institutions, research organizations or government agencies; each has four or five mentees on average.

The consortium follows a holistic approach, taking into account the pressures on women in a traditional culture and encouraging career–life balance through planning and coordination. It is developing scientists' skills in grant writing, leadership, ethics, research quality and time management. **Rose G. F. Leke, S. Kwedi Nolna** *University of Yaoundé, Cameroon. roseleke@yahoo.com*

Refereed science to guide action on EDCs

In a non-peer-reviewed venue (*Nature* **535**, 355; 2016), Daniel Dietrich *et al.* put forward apparently unsubstantiated arguments that in effect dismiss thousands of peer-reviewed academic studies and rigorous evaluations of endocrine-disrupting chemicals (EDCs) by independent scientists and organizations such as the World Health Organization (WHO), the United Nations Environment Programme, the Endocrine Society and the International Federation of

Gynecology and Obstetrics (see, for example, go.nature.com/2adgma2).

Identification of hazards associated with EDCs relies on randomized mechanistic studies in animals and observational epidemiological studies in humans. Randomized trials of direct chemical exposures in humans present serious ethical and other challenges. Using approaches developed by the WHO and the Intergovernmental Panel on Climate Change to account for the totality of the laboratory and human evidence and assess the strength of the evidence, the costs of continued inaction on EDCs are estimated to be more than €150 billion (US\$167 billion) annually (L. Trasande *et al. J. Clin. Endocrinol. Metab.* **100**, 1245–1255; 2015).

Proactive prevention measures backed by strong scientific criteria are therefore needed to prevent disease and disability across the life course. Regulatory decision-making should rely on peer-reviewed research to evaluate EDCs (L. Trasande *et al. J. Epidemiol. Comm. Health* <http://doi.org/bm5q>; 2016). Many of us have productively worked to achieve scientific consensus for the proposed criteria for EDCs in Europe (see go.nature.com/2awvrsn). In our view, Dietrich *et al.* do nothing to advance the debate or scientific knowledge on this important human health issue.

Leonardo Trasande *NYU School of Medicine, New York City, USA. leonardo.trasande@nyumc.org* Supported by 25 signatories (see go.nature.com/2aotwpi for full list).

CORRECTION

The print version of the Correspondence by M. W. Hayward *et al.* (*Nature* **534**, 475; 2016) incorrectly stated that 16 rhinos will be moved from Africa to Australia this year; in fact, it will be 6.

HEART DISEASE

Death-defying plaque cells

Dead cells are usually removed through their ingestion and destruction by other cells. A study of plaque deposits in arteries shows that dying cells in plaques display a 'don't-eat-me' signal that blocks their removal. [SEE LETTER P.86](#)

IRA TABAS

Heart attacks and strokes, which are leading causes of death worldwide¹, begin with a process called atherosclerosis, in which plaques — accumulations of lipids, cells, extracellular matrix and cellular debris — occur in certain areas of arteries. Although most people's arteries contain many such plaques, only a small percentage will cause disease². On page 86, Kojima *et al.*³ provide a plausible mechanism that could explain why some plaques become clinically dangerous.

A key feature of clinically dangerous ('vulnerable') plaques is a structure called the necrotic core, which contains dead cells that have undergone a type of cell death known as necrosis. The necrotic core is inflamed and has a thinning fibrous cap that covers the plaque and separates it from the central lumen of the artery² (Fig. 1). When the cap ruptures or erodes, the necrotic material becomes exposed to circulating blood-cell fragments called platelets that are necessary for blood clotting. This exposure results in platelet aggregation

(thrombus), which may block the blood vessel and thereby cause a heart attack or stroke by depriving the heart or brain of oxygen. The necrotic core, which harbours inflammatory cellular debris, promotes cap disruption by contributing to the degradation of the cap's structural protein, collagen, and by creating physical stress on the cap⁴. Understanding how the necrotic core develops is an urgent goal in heart-disease research.

To determine how dying cells in plaques undergo necrosis, it is necessary to understand how the body normally prevents necrotic cell death. Billions of cells in the body die every day through a process called apoptosis, which initially prevents cell-membrane rupture and leakage of inflammatory cellular contents. Apoptotic cells are rapidly and safely removed by an evolutionarily conserved process called efferocytosis, in which the apoptotic cell is internalized and destroyed by an engulfing cell, called a phagocyte, before membrane rupture occurs.

Efferocytosis requires signalling between the dying cell and the phagocyte: factors

produced by the apoptotic cell promote the migration of phagocytes towards apoptotic cells, and 'eat-me' recognition markers on the surface of apoptotic cells are recognized by receptors on phagocytes⁵. As a fail-safe mechanism, healthy living cells often express 'don't-eat-me' molecules on their cell surface that signal to block phagocytes from internalizing a live cell. The CD47 protein is an example of a don't-eat-me molecule that signals through the SIRPα receptor protein on phagocytes to inhibit apoptotic-cell engulfment⁵.

What goes wrong in vulnerable plaques? Studies have shown that efferocytosis is defective in 'advanced' human plaques that have not yet reached the vulnerable stage⁶, and experiments using genetically engineered mice⁴ have demonstrated a causal relationship between defective efferocytosis and plaque necrosis. Thus, in advanced plaques, uncleared apoptotic cells eventually become leaky, resulting in a process called secondary necrosis.

Why does efferocytosis become defective in advanced atherosclerosis? Kojima and colleagues provide a plausible mechanism. They

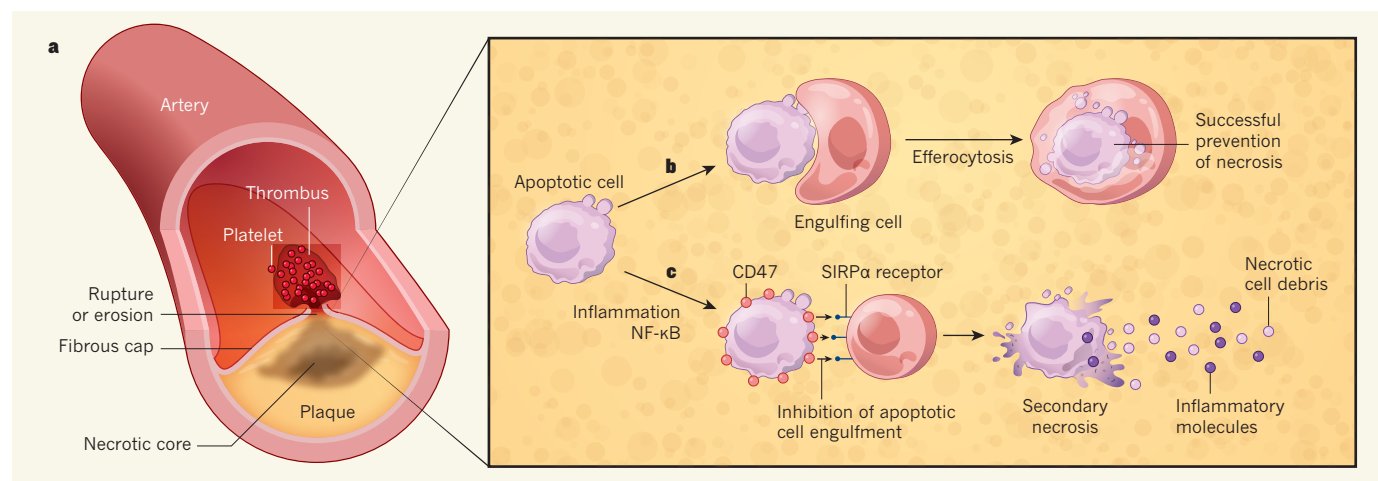


Figure 1 | Defective removal of dead cells can contribute to clinically dangerous atherosclerotic plaques. **a**, Many clinically dangerous plaques contain a structure called the necrotic core, characterized by inflammation and necrotic cell death. In atherosclerosis, if the fibrous cap covering the plaque ruptures or erodes, release of material from the necrotic core can trigger platelet aggregation (known as a thrombus) and arterial blockage, which may result in heart attack or stroke. Understanding how plaques develop to a necrotic state is a key question. **b**, Plaque cells undergo a non-inflammatory type of cell death called apoptosis. In asymptomatic

non-necrotic plaques, rapid removal of apoptotic cells by engulfing cells — a process known as efferocytosis — prevents necrosis. **c**, Kojima *et al.*³ found that the inflammatory conditions of advanced atherosclerosis lead to persistent expression of the protein marker CD47 on plaque cells through the inflammatory-signalling mediator NF-κB. When these cells become apoptotic, CD47 sends a signal through the SIRPα receptor on the engulfing cell to block engulfment. The unengulfed cells undergo a type of cell death called secondary necrosis, leading to the release of inflammatory molecules and the formation of necrotic cores from the cell debris.

made the surprising finding that in histological sections from human and mouse plaques, unengulfed dying macrophage and vascular smooth muscle cells display the don't-eat-me signal CD47 on their surface. In a mouse model of atherosclerosis, the authors found that infusion of an antibody that blocks CD47 improved efferocytosis in the plaque and lessened formation of the necrotic core. On the basis of an *in vitro* model, they suggest that CD47 is transcriptionally induced by NF- κ B, which orchestrates inflammatory programs in cells, including plaque cells. Defective phagocytic clearance of cells that die by another mechanism — an enzyme-triggered necrotic process called primary necrosis — may also contribute to the formation of the necrotic core⁷, and here too the problem could involve abnormal expression of CD47 (ref. 8).

The complex nature of both atherosclerosis and efferocytosis suggests that multiple mechanisms cause defective efferocytosis as plaques progress⁴. Workers from Kojima and colleagues' laboratory previously showed⁹ that dead cells in the plaque show a deficit in expression of the eat-me signal calreticulin protein. Moreover, the MerTK receptor protein present on phagocytic macrophages, which mediates efferocytosis in advanced plaques, undergoes degradation in the same type of inflammatory condition in atherosclerosis that Kojima and colleagues suggest leads to expression of CD47. The protease enzyme ADAM17 activates tumour necrosis factor- α (TNF- α), which induces CD47 in vascular smooth muscle cells, and ADAM17 also destroys MerTK¹⁰. Both ADAM17 activation and cleavage of MerTK have been implicated in the progression of human plaques towards a clinically dangerous state¹¹.

How might our knowledge of defective efferocytosis in general, and the insights gained from the work of Kojima and colleagues in particular, lead to future therapies to block the formation of dangerous plaques? Treatment with anti-TNF- α antibodies would block CD47 induction, and this strategy has been successful in debilitating autoimmune diseases for which TNF- α is a dominant trigger, notably rheumatoid arthritis. However, in atherosclerosis, it is probable that inflammation occurs through multiple pathways. Another concern is that anti-TNF- α treatment can compromise immune defences, which would challenge its long-term use as a preventive therapy in mostly asymptomatic people at risk of acute heart disease¹².

Treatment with anti-CD47 antibodies, which is being tested as a cancer treatment in early clinical trials¹³, presents other challenges. CD47 is used by red blood cells to prevent their premature engulfment before cell senescence, and a major adverse effect of anti-CD47 therapy is anaemia¹⁴ (a decrease in the number of red blood cells). Moreover, CD47 has roles in cell adhesion and migration, so its inhibition

might cause adverse effects related to these functions in processes such as blood-vessel formation and immune defence.

Another therapeutic strategy is based on the observation that many processes that generate vulnerable plaques, including inefficient efferocytosis, can be caused by defects in a biological program known as resolution of inflammation, which normally terminates an inflammatory response when it is no longer needed, and initiates tissue repair. Administration of compounds that mediate this resolution program has proved beneficial in many preclinical models of resolution-defective diseases¹⁵. For example, such treatment can improve efferocytosis and suppress plaque necrosis in advanced atherosclerosis¹⁶. Moreover, resolution-mediator therapy may actually boost host defence¹⁵, and this approach is now being tested in early clinical trials targeting chronic inflammatory conditions¹⁷. These and other future developments based on work such as that of Kojima and colleagues may some day provide a safe way to keep the plaques in our arteries from becoming clinically dangerous. ■

Ira Tabas is in the Departments of Medicine, Pathology and Cell Biology, and Physiology,

Columbia University School of Medicine, New York, New York 10032, USA.
e-mail: iat1@columbia.edu

1. World Health Organization. <http://www.who.int/mediacentre/factsheets/fs310/en> (2014).
2. Virmani, R., Burke, A. P., Kolodgie, F. D. & Farb, A. J. *Interv. Cardiol.* **15**, 439–446 (2002).
3. Kojima, Y. *et al.* *Nature* **536**, 86–90 (2016).
4. Thorp, E., Subramanian, M. & Tabas, I. *Eur. J. Immunol.* **41**, 2515–2518 (2011).
5. Arandjelovic, S. & Ravichandran, K. S. *Nature Immunol.* **16**, 907–917 (2015).
6. Schrijvers, D. M., De Meyer, G. R., Kockx, M. M., Herman, A. G. & Martinet, W. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1256–1261 (2005).
7. Lin, J. *et al.* *Cell Rep.* **3**, 200–210 (2013).
8. Greenlee-Wacker, M. C. *et al.* *J. Immunol.* **192**, 4709–4717 (2014).
9. Kojima, Y. *et al.* *J. Clin. Invest.* **124**, 1083–1097 (2014).
10. Thorp, E. *et al.* *J. Biol. Chem.* **286**, 33335–33344 (2011).
11. Garbin, U. *et al.* *Cardiovasc. Res.* **97**, 125–133 (2013).
12. Tabas, I. & Glass, C. K. *Science* **339**, 166–172 (2013).
13. <https://clinicaltrials.gov/ct2/show/NCT02678338>
14. Liu, J. *et al.* *PLoS One* **10**, e0137345 (2015).
15. Serhan, C. N. *Nature* **510**, 92–101 (2014).
16. Fredman, G. *et al.* *Sci. Transl. Med.* **7**, 275ra20 (2015).
17. <http://www.businesswire.com/news/home/20090824005320/en/Resolvex-Announces-Positive-Data-Phase-2-Clinical>

This article was published online on 20 July 2016.

SYNTHETIC BIOLOGY

Bacteria synchronized for drug delivery

A synthetic genetic circuit that mimics the quorum-sensing systems used by bacterial populations to coordinate gene expression enables bacteria to deliver drugs to mouse tumours in repeated and synchronized cycles. [SEE LETTER P.81](#)

SHIBIN ZHOU

Humans and bacteria have a long history of parasitic and symbiotic relationships. Now, Din *et al.*¹ exploit a relationship between bacteria and diseased human tissue for a therapeutic purpose. On page 81, the authors outline a system in which engineered bacteria acting as drug-delivery vehicles simultaneously break down, releasing an antitumour drug in synchronized cycles to maximize delivery efficiency and minimize toxicity.

In the body, some niches for bacteria — such as the anaerobic lumen of the intestines — have low oxygen levels. Similar conditions are found in solid tumours because of increased oxygen demand owing to highly proliferative tumour cells and insufficient blood supply owing to a structurally and functionally abnormal tumour vasculature². The hypoxic areas in a tumour are relatively protected from attacks by the body's

immune system, further facilitating bacterial colonization and growth³.

The idea of using bacteria to fight cancer has been around for more than a century. In 1891, surgeon William B. Coley infected patients with *Streptococcus* bacteria in an attempt to activate the immune system to fight cancer⁴. The method was controversial because of inconsistent efficacy and the toxicity of streptococcal infection. But the idea resurfaced later, when more was known about the tumour microenvironment and genetic-engineering tools had emerged, raising the hope that more-potent and less-toxic (attenuated) bacterial strains could be generated. Several bacterial strains have now been developed as agents for cancer therapy and they are showing promising effects in experimental models⁵.

Bacteria can destroy diseased tissue by competing for nutrients, secreting toxins and eliciting host immune responses. They

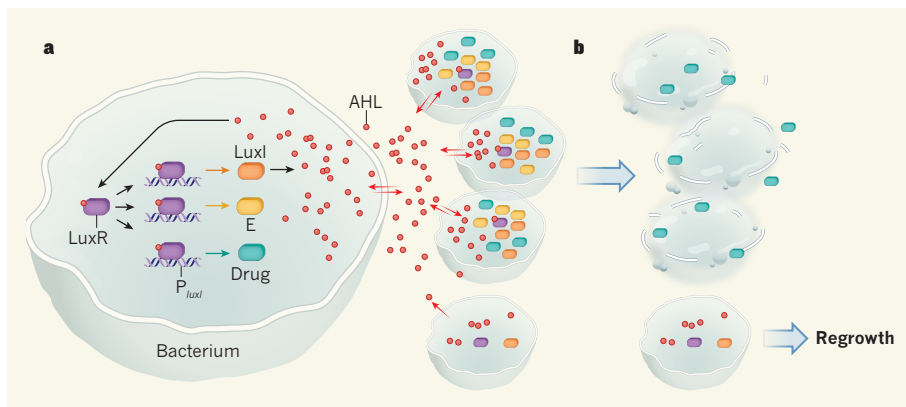


Figure 1 | Synchronized cyclical lysis and drug release. Din *et al.*¹ constructed a genetic circuit in tumour-targeting bacteria that mediates drug production and release through cell lysis in synchronized, repeated cycles. **a**, Binding of the signalling molecule AHL to its receptor protein LuxR leads to their subsequent interaction with, and activation of, the promoter DNA sequence P_{luxI} . This promoter drives expression of the gene that encodes the enzyme LuxI, which catalyses AHL synthesis, thus generating a feed-forward loop. The promoter also drives expression of genes that encode a bacterial toxin (drug) to kill cancer cells, and protein E, which releases the drug through bacterial-cell lysis. AHL can diffuse freely in and out of cells (red arrows). When the density of the bacterial population is low, AHL primarily diffuses out of the cell and the circuit is not active. When population density increases, intracellular AHL accumulates, reaching a threshold concentration that activates the circuit in most cells. **b**, The cells lyse in synchrony, releasing a burst of the drug. The few bacteria that survive kick off another cycle.

can also be genetically engineered to have extra antitumour activity. Compared with viruses, which have also been used in cancer treatments, bacteria have substantially greater capacity to carry non-native DNA. It is routine practice in molecular biology to introduce DNA stretches of several kilobases into a bacterial host, and bacterial artificial chromosomes greater than 300 kb can be transferred to, and maintained in, *Escherichia coli*⁶. In principle, therefore, bacteria can serve as efficient drug-delivery vehicles, carrying genetic circuits that encode and regulate therapeutic payloads.

An ideal drug-delivery system for cancer therapy should deliver the substance selectively to the tumour to minimize harm to healthy tissues, and should release the drug in a controlled manner. In their attempt to develop such a system, Din *et al.* focused on quorum-sensing circuits, which enable bacteria to communicate with one another, regulating gene expression in response to changes in population density.

In a previous study⁷, the authors of the current work used a synthetic-biology approach to construct a quorum-sensing genetic circuit in *E. coli*. Three components — LuxI, LuxR and acyl-homoserine lactone (AHL) — have crucial roles in this circuit. The enzyme LuxI catalyses synthesis of AHL molecules, and LuxR is an AHL receptor protein that activates a quorum-sensing transcriptional program. When bacterial population density is low, LuxI is expressed at a basal level. The AHL molecules synthesized as a result of LuxI expression do not accumulate in the cell, but instead rapidly diffuse out and become diluted in the extracellular environment. When population density rises, AHL accumulates in the

cell owing to the lowered diffusion gradient across the cell membrane. On reaching a threshold concentration in the cell, AHL molecules bind to LuxR. In turn, LuxR activates a promoter DNA sequence called P_{luxI} , which drives expression of target genes. Notably, because AHL can diffuse across the cell membrane, it reaches similar concentrations in all bacterial cells in the growing population, ensuring synchronized execution of the gene-expression program.

In the current study, Din *et al.* created a version of this genetic circuit that controlled synchronized and cyclical release of a bacterial toxin in attenuated *Salmonella enterica* serovar Typhimurium strains. In this system, P_{luxI} promotes expression of genes encoding four components — LuxI; the drug; a fluorescent protein to enable monitoring of population dynamics and drug release; and protein E, a lysis protein from a bacterial virus called ϕ X174 (Fig. 1). When the bacterial population reaches the critical density threshold, this P_{luxI} -driven transcriptional program is turned on in almost all cells, leading to drug production and its subsequent release owing to breakdown of bacterial cells through lysis. A few outliers in the population survive and repopulate the niche. The result is periodic bacterial lysis and drug delivery. To demonstrate efficacy, Din and colleagues treated tumour-bearing mice with the engineered bacteria and showed that the bacteria exhibit synchronized cyclical population dynamics, and confer some therapeutic benefits either when administered alone or in combination with chemotherapy.

The authors did not directly compare the efficacy of their bacteria with that of microbes engineered in the conventional way to

continuously secrete the therapeutic protein. Regardless, the new bacteria are notably different from the conventional ones. First, drug delivery is achieved through the simultaneous lysis of the entire population, rather than through continuous secretion by proliferating individuals. Second, the periodic lysis serves as a safety mechanism, because keeping the bacterial population to a defined size minimizes the risk of an adverse systemic inflammatory response that might harm the patient.

Despite these features, bacteria alone (whether engineered or not), are unlikely to eradicate tumours^{1,5}. In the current study, treatment of mice with the engineered microbes in combination with chemotherapy did not destroy the tumour; instead tumours shrank for 18 days, after which regrowth occurred. A curative therapeutic approach would most likely involve further improvements to engineered bacteria or using bacteria in combination with immunotherapy or other, more-powerful anticancer agents.

Cyclical drug release could be more useful for treating people who have diseases that require periodic dosing, such as diabetes and high blood pressure. To treat non-cancerous diseases, perhaps natural niches could be targeted for cyclical bacterial colonization. Alternatively, an implantable, semipermeable cassette that can be traversed by proteins and small molecules but not by bacteria could be developed to host the engineered microbes. One challenge to using periodically lysing bacteria to treat non-cancerous diseases that require long-term treatment is that the bacterial-degradation by-products released in each lysis cycle might be absorbed into the blood and build up, causing toxic systemic effects. Choosing less-toxic bacterial strains or creating attenuated strains (for example, by deleting the *msbB* gene, which is involved in making endotoxin⁸) could overcome the problem. ■

Shibin Zhou is in the Sidney Kimmel Comprehensive Cancer Center and the Ludwig Center for Cancer Genetics and Therapeutics at the Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. e-mail: sbzhou@jhmi.edu

1. Din, M. O. *et al.* *Nature* **536**, 81–85 (2016).
2. Vaupel, P., Mayer, A. & Höckel, M. *Methods Enzymol.* **381**, 335–354 (2004).
3. Yu, Y. A. *et al.* *Nature Biotechnol.* **22**, 313–320 (2004).
4. McCarthy, E. F. *Iowa Orthop. J.* **26**, 154–158 (2006).
5. Forbes, N. S. *Nature Rev. Cancer* **10**, 785–794 (2010).
6. Shizuya, H. *et al.* *Proc. Natl Acad. Sci. USA* **89**, 8794–8797 (1992).
7. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. *Nature* **463**, 326–330 (2010).
8. Low, K. B. *et al.* *Nature Biotechnol.* **17**, 37–41 (1999).

The author declares competing financial interests. See online article for details.

This article was published online on 20 July 2016.

ATOMIC PHYSICS

A milestone in quantum computing

Quantum computers require many quantum bits to perform complex calculations, but devices with more than a few bits are difficult to program. A device based on five atomic quantum bits shows a way forward. [SEE LETTER P.63](#)

STEPHEN D. BARTLETT

Quantum-savvy entrepreneurs are already bringing the first quantum computer processors out of the physics laboratory and onto the market. But these devices are mostly designed to perform just one function and cannot be programmed to run different algorithms. It would therefore be advantageous to build a fully fledged quantum computer that could be programmed to run anything we might want. In particular, it might execute the complex quantum algorithms that researchers think will solve today's intractable problems in quantum chemistry, materials science and data security. On page 63, Debnath *et al.*¹ present a small but fully programmable quantum computer consisting of five quantum bits (qubits), and they demonstrate its functionality by running several simple quantum algorithms.

Debnath and colleagues' computer is based on one of the oldest and most developed quantum architectures, which dates back to a design² proposed by physicists Ignacio Cirac and Peter Zoller in 1995. In this design, the computer's qubits are individual atomic ions that are trapped in a line using magnetic fields and manipulated with lasers (Fig. 1). The trapped ions behave like a tiny crystal, and precisely controlled vibrations along this line can cause the ions to become 'entangled'. Entanglement is a key ingredient of quantum computing whereby two or more particles

share a common state, such that each particle can no longer be described independently. Unlike most other quantum-computer architectures, the operations used to entangle the particles are not restricted only to neighbouring qubits.

Decades of research into precision metrology, such as the development of atomic clocks, now allow the quantum electronic states of trapped ions to be manipulated at an exquisite level of control and stability. Debnath and collaborators took advantage of this work and have also made several improvements to Cirac and Zoller's design, including the ability to target each ion (in this case, five ytterbium ions) individually with optical lasers. The net result is an elementary quantum processor in which every basic operation — initializing the states of the qubits, transforming them, entangling any pair of ions, and reading out the ions' quantum state — can be performed with errors occurring less than 2% of the time.

The accuracy of the authors' quantum processor allowed them to develop preset quantum logic gates that enact a desired sequence of laser pulses to generate an elementary component of a quantum circuit. In addition, they built a compiler that can take a quantum program — an algorithm designed to exploit some aspect of quantum mechanics to solve a mathematical problem — and determine how to operate the hardware to run the program.

The problems that can be solved by a small

computer with only five qubits are limited — they could be solved faster with even the most sluggish conventional laptop. But nonetheless, running simple algorithms can yield valuable information about the performance of the quantum processor as a whole, even when the outcome of the algorithm is already known. Why? A key concern for quantum architects is that qubits may seem to operate well when viewed individually, but can fail in unknown ways when required to work in tandem with many other qubits as part of a complex system. Simple algorithms are therefore used as a benchmark to see how several qubits function when combined in a larger circuit.

Debnath *et al.* demonstrate several algorithms. These include the Deutsch–Jozsa³ and Bernstein–Vazirani⁴ algorithms, which both use quantum effects to perform a mathematical calculation in a single step, whereas a conventional computer would require several operations. They also demonstrate a quantum Fourier transform^{5,6}, which is a key component of many of the heftier quantum algorithms, such as those used to break encryption. In all of these demonstrations, the resulting error rate is consistent with the authors' observations of how their qubits work in isolation, showing that the qubits can be used together in more-sophisticated algorithms in the future.

There is still a long way to go before quantum computers can reach their full potential. For the trapped-ion architecture explored here, researchers have already hit the limit of the number of ions that can be placed in a line in a single trap — around a dozen⁷. The future of this field is believed to involve either joining many such traps together using optical quantum couplers, or shuttling ions between interaction zones in microfabricated traps that have a 2D layout⁸. The latter approach also offers the tantalizing possibility of low error rates for basic logic operations, perhaps even just one error in every thousand operations — a figure commonly thought to be the highest error rate that a large-scale quantum computer could tolerate. Research in these directions has been encouraging, but it may be a while before these scalable approaches can reproduce even the five-qubit results demonstrated by Debnath and colleagues' quantum computer.

Trapped-ion quantum architectures are not the only game in town. A range of other solid-state, atomic and optical quantum systems each have different advantages for quantum computing. Notably, an approach using qubits built of superconducting circuitry — considered the dark horse of quantum computing research only a decade ago — has shown enormous recent success^{9,10}. Not only can superconducting technologies now compete with the phenomenal precision that has been shown with trapped ions, but they can also operate at much higher speeds and may have a clearer pathway to being scaled up.

A programmable five-qubit quantum

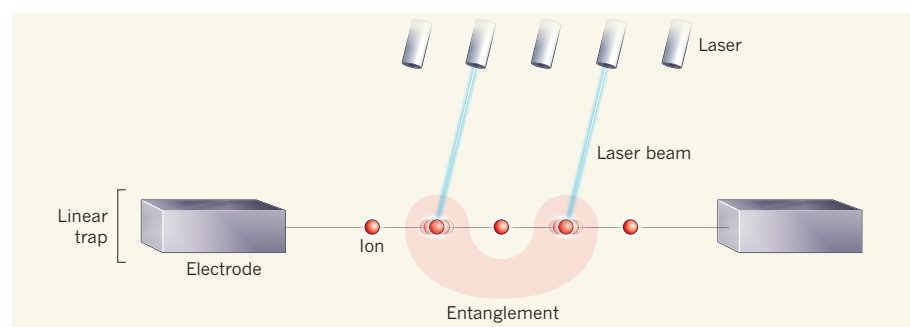


Figure 1 | Design of an atomic quantum computer. Debnath *et al.*¹ have constructed a quantum computer that is based on an earlier design². The computer's quantum bits (qubits) are individual atomic ions that are trapped in a line using magnetic fields and electrodes. The ions are carefully manipulated by lasers so as to vibrate. The vibration of two ions allows them to become 'entangled', enabling quantum computations to be performed.

computer built using superconducting circuits has now also been demonstrated¹¹, and has similar capabilities to Debnath and colleagues' device. Both the superconducting-circuit and ion-trap approaches seem to be capable of being scaled up to larger devices that have more quantum bits. The next challenge for all of these technologies is to demonstrate that quantum error correction can bring error rates down to negligible levels. ■

Stephen D. Bartlett is at the Centre for Engineered Quantum Systems and the School of Physics, University of Sydney, Sydney, New South Wales 2006, Australia.
e-mail: stephen.bartlett@sydney.edu.au

1. Debnath, S. *et al.* *Nature* **536**, 63–66 (2016).
2. Cirac, J. I. & Zoller, P. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
3. Deutsch, D. & Jozsa, R. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).

4. Bernstein, E. & Vazirani, U. *SIAM J. Comput.* **26**, 1411–1473 (1997).
5. Shor, P. *SIAM J. Comput.* **26**, 1484–1509 (1997).
6. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 2nd edn (Cambridge Univ. Press, 2002).
7. Monz, T. *et al.* *Phys. Rev. Lett.* **106**, 130506 (2011).
8. Kielpinski, D., Monroe, C. & Wineland, D. J. *Nature* **417**, 709–711 (2002).
9. Barends, R. *et al.* *Nature* **508**, 500–503 (2014).
10. Chow, J. M. *et al.* *Nature Commun.* **5**, 4015 (2014).
11. Takita, M. *et al.* Preprint at <https://arxiv.org/abs/1605.01351> (2016).

BIOMEDICAL SCIENCE

Protection for anaesthetized mice

A cognition-enhancing drug called CX546 prevents the neurodegenerative effects of repeated anaesthesia in infant mice by promoting neuronal changes associated with learning and by protecting neurons from death.

LAURA CORNELISSEN & CHARLES BERDE

Millions of children have surgery under general anaesthesia each year. Studies of infant animals show that neurodegeneration and long-term neurobehavioural impairments arise when general anaesthesia is used at crucial periods of brain development, especially following high doses and prolonged exposures to anaesthetics¹. Studies in humans have been controversial — some have reinforced these findings, particularly among infants who have had multiple anaesthetics and surgeries², whereas a case-control study³ and an interim analysis⁴ of a recent randomized trial have been more reassuring. Writing in *Science Translational Medicine*, Huang *et al.*⁵ report that a drug called

CX546 confers neuroprotection in infant mice that are repeatedly exposed to the anaesthetic molecule ketamine.

Ketamine is thought to act predominantly by blocking signalling through NMDA receptor proteins⁶, which are activated by the excitatory neurotransmitter molecule glutamate. The authors examined the effects of ketamine anaesthesia on neuronal activity in the brains of infant mice. *In vivo* imaging experiments revealed that neuronal activity decreased during post-anaesthesia recovery in treated animals compared to untreated animals. Analysis of proteins at the synaptic junctions between neurons showed that, after repeated anaesthesia, expression of NMDA receptors was reduced in adulthood, as was expression of another class of

glutamate receptor proteins, AMPA receptors.

CX546 is part of a group of cognition-enhancing drugs called AMPAkinases that assist excitatory neurotransmission through AMPA receptors. Huang and colleagues found that CX546 prevented ketamine-induced death of brain neurons in infant mice, restored the expression of AMPA and NMDA receptors, and preserved neuronal activity in vulnerable brain regions. Moreover, the drug improved neurobehavioural outcomes, for example by rescuing the learning deficits that are associated with repeated ketamine anaesthesia (Fig. 1). Finally, the authors showed that CX546 partially rescued the remodelling of dendritic spines — tiny neuronal structures whose formation and elimination are crucial for processes such as learning. These structures cannot be correctly remodelled following repeated ketamine anaesthesia^{7,8}.

Glutamate is the most prominent excitatory neurotransmitter in the central nervous system, and plays a crucial part in the processes of neural-circuit strengthening (through sustained activity) and elimination (through weak activity). CX546 enhances glutamate-mediated neurotransmission and strengthens circuits by increasing neuronal activity. This is probably how the drug provides neuroprotection when it is given immediately after the periods of low neuronal activity that follow repeated ketamine anaesthesia.

Several AMPAkinases are already in clinical trials in adults as treatments for a range of conditions, including Parkinson's disease, schizophrenia and autism⁹. Huang *et al.* speculate that CX546 might hold promise as a therapy to prevent neuronal defects in human infants undergoing surgery and anaesthesia.

Widely varying recommendations and mitigation strategies have been proposed in response to the debate around anaesthetic-induced neurotoxicity in human infants¹⁰. In 2012, a public-private collaboration between the US Food and Drug Administration and the International Anesthesia Research Society, called SmartTots, recommended delaying elective surgery that uses general anaesthesia until patients are at least three years old whenever possible¹¹. Subsequently, these authorities amended their recommendations to advocate balancing the risks and benefits to guide individual treatment decisions¹².

For many types of paediatric surgery, such

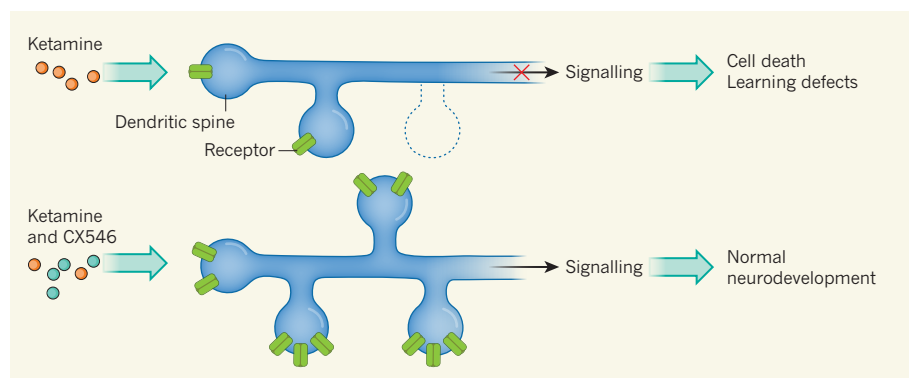


Figure 1 | Combating ketamine. **a**, In infant mice, repeated exposure to anaesthesia using the molecule ketamine inhibits neuronal signalling mediated by AMPA and NMDA receptor proteins, and prevents remodelling of tiny neuronal structures called dendritic spines (dashed line indicates a spine whose remodelling was prevented). This causes neuronal death and weakening of neural circuits, leading to defects in learning and motor performance. **b**, Huang *et al.*⁵ report that a drug called CX546, which increases AMPA and NMDA expression and so promotes excitatory neurotransmission, can rescue these defects, preserving normal neurodevelopment.

a delay is either not feasible or would probably do more harm than good. For instance, delayed repair of congenital heart defects could cause death or neurological deficits — a much greater risk than the potential consequences of exposure to a general anaesthetic. Similarly, many head and neck procedures performed during infancy and early childhood foster optimal neurodevelopment by correcting impairments in hearing, vision, speech or feeding, or by removing airway obstructions. In these cases, avoiding general anaesthesia is simply not practical. However, local or regional anaesthesia can sometimes be used to reduce dose requirements for general anaesthetics, to provide postoperative pain relief, or, in selected cases, to act as a primary anaesthetic⁴.

Alternative anaesthetic agents such as dexmedetomidine and xenon are currently under investigation for use in the clinic, on the basis of animal data¹³ suggesting that they cause less neurotoxicity than ketamine or the widely used inhalation anaesthetics sevoflurane and isoflurane. Dexmedetomidine and xenon are not complete anaesthetics by themselves at clinically achievable doses, and there are some practical barriers to implementation. Nonetheless, both hold some promise for drastically reducing the doses of conventional anaesthetics that are required to maintain a general anaesthetic state.

Several questions should be addressed before testing CX546 in clinical trials. First, is the neuroprotective action of CX546 specific for ketamine, or can the drug also protect against anaesthetics that act on different neuronal circuits? Second, how crucial is the timing of CX546 administration for its neuroprotective effect? Huang *et al.* gave CX546 to infant mice after anaesthesia, but delivering it during anaesthesia and surgery might change the dose requirements.

Third, because AMPAkinases have been shown to stimulate respiration¹⁴, it is possible that the neuroprotective benefit of CX546 is partly due to respiratory stimulation. This in turn remedies the low levels of oxygen and high levels of carbon dioxide in the blood that could be induced by ketamine and other anaesthetics. Huang *et al.* did not analyse respiration during or after anaesthesia, so this remains to be investigated. Finally, the potential adverse effects of exposing brains to CX546 during crucial periods of their development must be assessed.

Preventive medicines have previously caused considerable harm¹⁵. For a medication to be beneficial in a preventive role, the average number of patients that can be treated before one person is harmed (the 'number needed to harm') must be extremely high, and the average number of patients who must be treated before one extra person benefits compared to the previous regime (the 'number needed to treat') must be comparatively low. Clinical-trial designs for a CX546 neuroprotection study in

infants undergoing surgery would face serious challenges¹⁶, and would require an extensive preclinical toxicology programme that tested infant animals from several species.

From the standpoints of ethics, risk–benefit and effect size, it might be appropriate to conduct a clinical trial among infants who are already having repeated or prolonged surgical procedures or who require long-term sedation. These infants have complex and varied medical conditions and a range of confounding factors that would make a prevention trial difficult, but not impossible. It would be unethical not to use a trial design that is randomized, double blind, prospective (one that studies subjects after enrolment, rather than retrospectively) and controlled^{15,16}.

Huang *et al.* are to be commended for an innovative study, which introduces a plausible, mechanism-based potential preventive treatment for anaesthetic-induced neurotoxicity in infants. Their work adds to our understanding of the mechanisms that underpin ketamine's activity, its effects and the potential interventions that could optimize neurodevelopment in infants undergoing surgery and anaesthesia. ■

Laura Cornelissen and Charles Berde
are in the Department of Anesthesiology,

DIABETES

Still a geneticist's nightmare

The largest DNA-sequencing study of type 2 diabetes conducted so far concludes that, contrary to expectation, low-frequency and rare genetic variants do not contribute significantly to disease risk. [SEE ARTICLE P.41](#)

STEPHEN S. RICH

Type 2 diabetes is a major cause of illness and death, particularly in people of African, Hispanic and Asian ancestry. Despite the indications of strong familial origins¹, it was not until 2012 that genetic variants associated with the disease were robustly established, thanks to a genome-wide association study (GWAS) that looked for common risk variants in more than 100,000 people from several ancestral populations². Nonetheless, only around 10% of the risk attributable to genetic factors has been identified. An obvious next approach is to interrogate the genome for infrequent and rare variants that could affect risk individually or in aggregate. On page 41, Fuchsberger *et al.*³ present a comprehensive evaluation of the role of rare and infrequent variants in the risk of developing type 2 diabetes.

The authors sequenced exomes, which

Perioperative & Pain Medicine, Boston Children's Hospital, Boston, Massachusetts 02115, USA, and in the Department of Anaesthesia, Harvard Medical School, Boston. e-mail: charles.berde@childrens.harvard.edu

1. Disma, N., Mondardini, M. C., Terrando, N., Absalom, A. R. & Bilotta, F. *Paediatr. Anaesth.* **26**, 6–36 (2016).
2. Wang, X., Xu, Z. & Miao, C.-H. *PLoS ONE* **9**, e85760 (2014).
3. Sun, L. S. *et al. J. Am. Med. Assoc.* **315**, 2312–2320 (2016).
4. Davidson, A. J. *et al. Lancet* **387**, 239–250 (2016).
5. Huang, L., Cichon, J., Ninan, I. & Yang, G. *Sci. Transl. Med.* **8**, 344ra85 (2016).
6. Brown, E. N., Purdon, P. L. & Van Dort, C. J. *Annu. Rev. Neurosci.* **34**, 601–628 (2011).
7. Hayashi, H., Dikkes, P. & Soriano, S. G. *Paediatr. Anaesth.* **12**, 770–774 (2002).
8. Vutsits, L., Gascon, E., Tassonyi, E. & Kiss, J. Z. *Toxicol. Sci.* **91**, 540–549 (2006).
9. Urban, K. R. & Gao, W.-J. *Front. Syst. Neurosci.* **8**, 38 (2014).
10. Rappaport, B. A., Suresh, S., Hertz, S., Evers, A. S. & Orser, B. A. *N. Engl. J. Med.* **372**, 796–797 (2015).
11. <http://smarttots.org/wp-content/uploads/2015/08/SmartTots-Case-Statement-2013.pdf>
12. <http://smarttots.org/consensus-statement-supplement>
13. Maze, M. *Can. J. Anaesth.* **63**, 212–226 (2016).
14. van der Schier, R., Roozkrans, M., van Velzen, M., Dahan, A. & Niesters, M. *F1000Prime Rep.* **6**, 79 (2014).
15. Sackett, D. L. *Can. Med. Assoc. J.* **167**, 363–364 (2002).
16. Davidson, A. J. *et al. Paediatr. Anaesth.* **25**, 447–452 (2015).

encompass protein-coding regions (about 1–2% of the human genome), from around 6,500 people with type 2 diabetes and 6,400 healthy controls, from 5 ancestry groups. They also sequenced whole genomes from some 1,300 people with the disease and 1,300 ancestry-matched controls. After rigorous quality control, the whole-genome sequences revealed approximately 27 million sequences or bases that varied between individuals.

Fuchsberger and colleagues found that 126 of these variants, each located in one of four genes, were significantly associated with an altered risk of type 2 diabetes. Two of the genes, *TCF7L2* and *ADCY5*, had been previously identified as containing commonly occurring variants associated with diabetes risk, and a risk-associated variant in a third, *CCND2*, had been found to occur at low frequency⁴. The final gene, *EML4*, contained a common variant that had not been

identified from the GWAS — but this discovery was not replicated in a larger data set.

Crucially, there was no significant evidence for rare or low-frequency disease-associated variants in regulatory elements, which modulate gene expression, or in coding sequences. Combining samples that had undergone both exome and whole-genome sequencing revealed only one more risk variant, in the gene *PAX4*; this variant had been previously detected in people from East Asia^{5,6}.

To broaden their search, the authors increased the number of subjects to around 90,000. They analysed the participants' DNA using a customized array — a tool that allowed the analysis of specific coding sequences in which diabetes-associated variation might arise, thereby generating fewer raw data than genome sequencing. In this expanded data set, they found another 18 common variants in 13 genes. Only one of these, in *MTMR3*, had not previously been identified by the GWAS. Fuchsberger and colleagues conclude that there is little evidence that rare or infrequent variants affect the risk of developing diabetes. Instead, the authors suggest that almost all significantly diabetes-associated variants are common in the population and have been previously detected by the GWAS.

If the contribution of common variants to the genetic risk of diabetes is relatively limited, and if there is little support for a contribution from rare and low-frequency variants, where are the culprits hiding? Once dubbed “a geneticist's nightmare”⁷, diabetes seems to be living up to its reputation. What are researchers missing? Could other genetic tools be applied?

One approach to improving the search would be to add more whole-genome sequences, in the hope that a larger population will enable smaller effects to be detected. Size does matter in genomic studies, but the current work suggests that increased sample size will not much increase the number of risk-associated genes or variants. However, the cost of increasing sample size might be surprisingly low. Several initiatives from the US National Institutes of Health, including the Trans-Omics for Precision Medicine program, the Centers for Common Disease Genomics project and the Personalized Medicine Initiative, will provide whole-genome sequences and associated data on phenotypes (the severity of a range of diabetes-associated traits) free of charge. Even if the expected yield of such analyses is low, finding a handful of rare variants — for example, those that confer loss of gene function, such as one found in *PCSK9* (ref. 8) — could have a major impact.

There are many alternatives to increasing sample size. These include focusing only on the variants that confer an increased risk of harmful phenotypes⁹, analysing diverse populations or those known to be at low risk of disease¹⁰, and considering other genetic ‘architectures’, such as whether the disease is caused by a modest contribution from many common variants,

rather than a large contribution from a few rare ones. Moreover, genetic risk should be considered in the context of the complex environmental risk factors with which it is inexorably intertwined. For example, being overweight does not always lead to diabetes, but fatty tissue increases insulin resistance, and abdominal fat increases the risk of diabetes more than does fat in hips and thighs. Physical activity not only controls weight, but also uses glucose as energy and increases cellular insulin sensitivity. An understanding of genetic risk factors is needed to elucidate the mechanisms that lead to such complex, variable phenotypes.

Many genes and variants, both common and rare, could influence the declining function of the β -cells that store and release insulin. However, relatively small sets of variants might be sufficient to elevate risk in the context of other genetic or environmental factors. There might be cassettes of variants unique to an individual — called private variants — that lead to diabetes only in certain conditions. In this scenario, the large case-control design is not necessarily optimal, because the averaging of effects will obscure crucial gene sets. Family-based designs¹¹ and efforts to further dissect the subtleties of each phenotype might be required to identify private variants⁹.

Historically, to obtain a grasp of risk-associated variants, genetic studies took advantage of extremes of disease presentation (such as mild or severe phenotypes, or early or late diagnosis¹²), or of the occurrence of disease in low-risk groups, or of populations in which private variants might contribute to disease risk. Although such studies have already been performed, it might be time to revisit these

approaches using whole-genome-sequence and refined phenotypic data.

The authors' study marks the first large-scale use of whole-genome sequencing data to tackle this incredibly complex disease. The conclusion that rare and infrequent variants have little effect on risk under this study design and in these populations is important. However, it may be that Fuchsberger *et al.* have eliminated only the rare variants identifiable from a case-control study. The genetics of type 2 diabetes might still be a nightmare, but nonetheless the search continues. ■

Stephen S. Rich is in the Center for Public Health Genomics and Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, Virginia 22908-0717, USA.
e-mail: ssr4n@virginia.edu

1. Rewers, M. & Hamman, R. F. in *Diabetes in America* 2nd edn, Ch. 9, 179–220 (NIH, 1995).
2. Morris, A. P. *et al. Nature Genet.* **44**, 981–990 (2012).
3. Fuchsberger, C. *et al. Nature* **536**, 41–47 (2016).
4. Steinthorsdottir, V. *et al. Nature Genet.* **46**, 294–298 (2014).
5. Cho, Y. S. *et al. Nature Genet.* **44**, 67–72 (2012).
6. Ma, R. C. W. *et al. Diabetologia* **56**, 1291–1305 (2013).
7. Neel, J. V. in *The Genetics of Diabetes Mellitus* (eds Creutzfeldt, W., Köbberling, J. & Neel, J. V.) 1–11 (Springer, 1976).
8. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. N. *Engl. J. Med.* **354**, 1264–1272 (2006).
9. Bergman, R. N. *et al. Diabetes* **52**, 2168–2174 (2003).
10. Zeggini, E., Gloyn, A. L. & Hansen, T. *Diabetologia* **59**, 938–941 (2016).
11. Blangero, J. & Kent, J. W. Jr in *Genetic and Molecular Aspects of Sport Performance* (eds Bouchard, C. & Hoffman, E. P.) 33–45 (Wiley-Blackwell, 2011).
12. Lange, L. A. *et al. Am. J. Hum. Genet.* **94**, 233–245 (2014).

This article was published online on 11 July 2016.

QUANTUM PHYSICS

Destruction of discrete charge

Electric charge is quantized in units of the electron's charge. An experiment explores the suppression of charge quantization caused by quantum fluctuations and supports a long-standing theory that explains this behaviour. SEE LETTER P.58

YULI V. NAZAROV

Because matter is constructed of elementary particles, the electric charge of any object is an integer multiple of the elementary charge, which is equivalent in size to the electron's charge. This concept is known as charge quantization. In 1913, charge quantization (and thus, the existence of elementary particles) was demonstrated by the physicist Robert Millikan, who measured the charges of single electrons in oil drops containing many

billions of particles¹. However, quantum physics predicts that charge quantization can be destroyed by tiny quantum fluctuations. On page 58, Jezouin *et al.*² describe an experiment to control these fluctuations. Their results suggest that the effect of the fluctuations on charge quantization can be explained through particle-like phenomena called Korshunov instantons³.

In the current age of nanoscience, charge quantization has enabled the manipulation of single electrons in nanostructures, with applications in metrology, sensing and

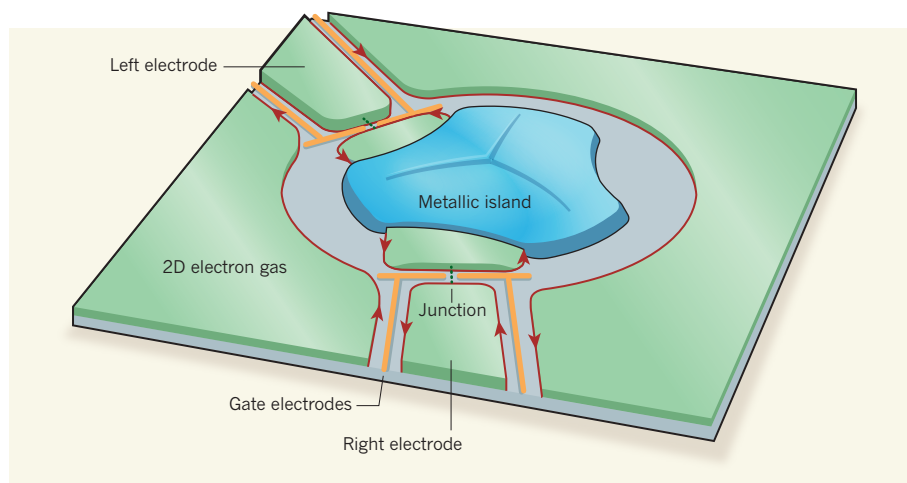


Figure 1 | A nanodevice to investigate charge quantization. Jezouin *et al.*² have studied the effect of quantum fluctuations on the quantization of electric charge. Their device consists of a micrometre-scale metallic island, which stores a discrete charge. The device is built on a semiconducting material, the surface of which supports a thin layer of electron gas (green). When a voltage is applied to two pairs of metallic 'gate' electrodes, the electrons of the gas are repelled, leading to regions in which the gas is depleted (grey) and allowing current to flow through the device (red arrows). The island is connected to two other electrodes (the left and right electrodes) through separate junctions. Electrons can travel through the junctions because of quantum tunnelling (dotted lines) with a probability that is controlled using the gate electrodes. Charge quantization is observed as Coulomb oscillations, a periodic dependence of the device's conductance on the voltage of one of the gate electrodes. The authors have measured the visibility of these oscillations against the voltage-independent component of conductance as the transmission probabilities of the junctions and the temperature are varied.

thermometry⁴. These nanostructures consist of conducting, metallic islands that store the single charges, and which are analogues of Millikan's oil drops. The islands are connected to electrodes and to each other by tunnel junctions — barriers between two conducting materials, such as thin layers of an insulator. Other 'gate' electrodes are used to control the discrete charges on the islands.

The most common manifestation of charge quantization in such nanostructures is the presence of Coulomb oscillations³: a periodic dependence of the structure's conductance on the voltage of one of the gate electrodes. However, if the islands are not well isolated, the conductance of the connecting junctions produces quantum fluctuations of charge that suppress the quantization, reducing the visibility of the Coulomb oscillations against the voltage-independent component of conductance.

The aim of Jezouin and colleagues' experiment² is to fully control these quantum fluctuations. Their set-up is remarkably simple: a single metallic island is connected to two electrodes through separate junctions (Fig. 1). Unlike in previous attempts to control the fluctuations (see, for example, refs 5–7), each junction has only one conduction channel (a pathway through which an electron can travel). Therefore, the conduction of each junction is directly related to its transmission probability, the probability that an electron will be transmitted through the junction.

The authors are able to vary and quantify the transmission probability of each junction. To achieve this level of control, the conduction channels are formed in a semiconductor

material, the surface of which supports a thin layer of electron gas. The authors' technological advance is to connect this two-dimensional gas to the metallic island.

Using this set-up, the authors measure the visibility of the Coulomb oscillations as two quantities — the transmission probabilities of the two junctions and the temperature — are varied. They find that the visibility of the oscillations is reduced if either of the junctions' transmission probabilities is increased. Furthermore, as the temperature is increased, thermal fluctuations result in an exponential suppression of the oscillations. What is the significance of these two 'scaling laws'?

Unlike classical fluctuations, which tend to destroy the order of a system, quantum fluctuations can actually generate alternative order. For example, large fluctuations in the position of a quantum particle imply that the particle's momentum has a precise value, whereas large fluctuations of momentum imply a precise position. Electric charge and magnetic flux are similarly connected in superconducting nanostructures⁸, so that increases in charge fluctuation convert charge quantization to flux quantization.

States of quantized flux do not exist in normal, non-superconducting metals. However, in 1987, the physicist Sergey Korshunov unexpectedly discovered that two flux quanta can be transferred between non-quantized flux states without the requirement for superconductivity, owing to the presence of particle-like phenomena called instantons⁹. Since then, the manifestations of these Korshunov instantons have been thoroughly investigated.

In 1999, the concept was extended to arbitrary transmission probabilities⁹, leading to two specific predictions that describe the influence of instantons on charge quantization.

These predictions are precisely the scaling laws observed by Jezouin and collaborators. The authors' results therefore strongly suggest that charge quantization over the whole range of fluctuation strength is governed by Korshunov instantons, and provide experimental evidence of these long-predicted 'particles'.

The Korshunov instanton is a close relative of another concealed 'particle', the leviton¹⁰. Levitons are generated when a carefully engineered voltage pulse, encompassing precisely two flux quanta, is applied to a collection of electrons in a nanojunction. This phenomenon was confirmed¹¹ experimentally in 2013. Whereas the leviton is an excitation of electron systems, the instanton is a property of the electronic ground state that, in the presence of electrostatic (Coulomb) interactions, governs charge quantization.

One puzzle left by the authors' experiment is why the scaling laws are consistent with the theoretical predictions for a single instanton. Theoretical considerations suggest that Jezouin and colleagues' system would involve configurations of many instantons, which would lead to more-complex scaling laws than those observed. Their results may imply that the theory that underlies such systems is simpler than it seems and has unknown symmetries that would cancel out many-instanton configurations. These theories could be tested using more-detailed measurements, especially of the island capacitance. One could also construct more-sophisticated set-ups that combine levitons and instantons. Greater control of these concealed particles could then be achieved by studying their interference.

Charge quantization is a simple concept; however, observing the effect of quantum fluctuations can lead to fascinating discoveries. Chasing exotic particles in nanostructures will help us to understand the complexity of quantum laws of electricity, with the hope of eventually applying this knowledge to quantum information processing. ■

Yuli V. Nazarov is at the Kavli Institute of Nanoscience Delft, 2628 CJ Delft, the Netherlands.
e-mail: y.v.nazarov@tudelft.nl

1. Millikan, R. A. *Phys. Rev.* **2**, 109–143 (1913).
2. Jezouin, S. *et al. Nature* **536**, 58–62 (2016).
3. Korshunov, S. E. *JETP Lett.* **45**, 434–436 (1987).
4. Likharev, K. K. *Proc. IEEE* **87**, 606–632 (1999).
5. Kouwenhoven, L. P. *et al. Z. Phys. B* **85**, 367–373 (1991).
6. Staring, A. A. M. *et al. Physica B* **175**, 226–230 (1991).
7. van der Vaart, N. C. *et al. Physica B* **189**, 99–110 (1993).
8. Devoret, M. H. & Schoelkopf, R. J. *Science* **339**, 1169–1174 (2013).
9. Nazarov, Y. V. *Phys. Rev. Lett.* **82**, 1245–1248 (1999).
10. Keeling, J., Klich, I. & Levitov, L. S. *Phys. Rev. Lett.* **97**, 116403 (2006).
11. Dubois, J. *et al. Nature* **502**, 659–663 (2013).

The genetic architecture of type 2 diabetes

A list of authors and affiliations appears in the online version of the paper

The genetic architecture of common traits, including the number, frequency, and effect sizes of inherited variants that contribute to individual risk, has been long debated. Genome-wide association studies have identified scores of common variants associated with type 2 diabetes, but in aggregate, these explain only a fraction of the heritability of this disease. Here, to test the hypothesis that lower-frequency variants explain much of the remainder, the GoT2D and T2D-GENES consortia performed whole-genome sequencing in 2,657 European individuals with and without diabetes, and exome sequencing in 12,940 individuals from five ancestry groups. To increase statistical power, we expanded the sample size via genotyping and imputation in a further 111,548 subjects. Variants associated with type 2 diabetes after sequencing were overwhelmingly common and most fell within regions previously identified by genome-wide association studies. Comprehensive enumeration of sequence variation is necessary to identify functional alleles that provide important clues to disease pathophysiology, but large-scale sequencing does not support the idea that lower-frequency variants have a major role in predisposition to type 2 diabetes.

There is compelling evidence that the individual risk of type 2 diabetes (T2D) is strongly influenced by genetic factors¹. Progress in characterizing the specific T2D-risk alleles responsible has been catalysed by the ability to perform genome-wide association studies (GWAS). Over the past decade, successive waves of T2D GWAS—featuring ever larger samples, progressively denser genotyping arrays supplemented by imputation against more complete reference panels, and richer ethnic diversity—have delivered more than 80 robust association signals^{2–8}. However, in these studies, the alleles interrogated for association were predominantly common (minor allele frequency (MAF) >5%), and with limited exceptions^{7,9}, the variants driving known association signals were also common, with individually modest impacts on T2D risk^{2–8,10}. Variation at known loci explains only a minority of observed T2D heritability^{2,3,11}.

Residual genetic variance is partly explained by a long tail of common variant signals of lesser effect². However, the contribution to T2D risk that is attributable to lower-frequency variants remains a matter of considerable debate, not least because of the relevance of disease architecture to clinical application¹¹.

Next-generation sequencing enables direct evaluation of the role of lower-frequency variants to disease risk^{7,12,13}. This paper describes the efforts of the coordinated, complementary strategies pursued by the Genetics of Type 2 Diabetes (GoT2D) and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia. GoT2D collected comprehensive genome-wide sequence data from 2,657 T2D cases and controls; T2D-GENES focused on exome sequence variation, assembling data (after inclusion of GoT2D exomes) from a multiethnic sample of 12,940 individuals. Both consortia used genotype data to expand the sample size available for association testing for a subset of the variants exposed by sequencing.

Analysis of genome-wide variation

The GoT2D consortium selected for whole-genome sequencing cases of type 2 diabetes (T2D) and ancestry-matched normoglycaemic controls from northern and central Europe (Methods and Supplementary Table 1). To increase power to identify low-frequency (0.5% < MAF < 5%) and rare (MAF < 0.5%) T2D variants with large effects, we preferentially identified individuals from the extremes of genetic risk (Methods). The genome sequence of 1,326 cases and 1,331 control individuals was determined through joint statistical analysis of low-coverage whole-genome sequence (~5×), deep-coverage

exome sequence (~82×), and array-based genotypes at 2.5 million single nucleotide variants (SNVs) (Extended Data Fig. 1 and Extended Data Table 1).

We detected, genotyped, and estimated haplotype phase for 26.7 million genetic variants (Extended Data Fig. 1 and Extended Data Table 2), including 1.5 million short insertion-deletion variants (indels) and 8,876 large deletions. Individual diploid genomes carried a mean of 3.30 million variants (range: 3.20 million–3.35 million), including 271,245 indels (262,201–327,077), and 669 (579–747) large deletions. These data include many variants not directly studied by previous GWAS, including all of the indels as well as 420,473 common and 2.4 million low-frequency SNVs that were poorly tagged ($r^2 \leq 0.30$)^{3,4} by genotype arrays. We estimate near-complete ascertainment (98.2%) of SNVs with minor allele counts of greater than 5 (MAF > 0.1%), and high accuracy (over 99.1%) at heterozygous genotypes (Methods and Fig. 1a). As half of the sequenced individuals were T2D cases, ascertainment was enhanced for any rare or low-frequency variants that substantially increase T2D risk (Fig. 1a). Specifically, we estimate ≥80% power to detect (at genome-wide significance, $\alpha = 5 \times 10^{-8}$) T2D risk variants with MAF ≥ 5% and odds ratio (OR) ≥ 1.87, or MAF ≥ 0.5% and OR ≥ 4.70 (Extended Data Fig. 2).

We tested all 26.7 million variants for T2D association by logistic regression assuming an additive genetic model (Supplementary Table 2). Analyses using a mixed-model framework to account for population structure and relatedness generated almost identical results. At genome-wide significance, 126 variants at four loci were associated with T2D (Fig. 1b). These included two previously reported common-variant loci (*TCF7L2* and *ADCY5*), a previously reported low-frequency variant in *CCND2* (ref. 7) (rs76895963, MAF = 2.6%, $P_{\text{seq}} = 4.2 \times 10^{-9}$), and a novel common-variant association near *EML4* (MAF = 34.8%, $P_{\text{seq}} = 1.0 \times 10^{-8}$). There was no significant evidence of association with T2D for sets of low-frequency or rare variants within coding regions, nor within specified non-coding regulatory elements (Methods).

Power to detect association with low-frequency and rare variants of modest effect is limited in a sample of 2,657 individuals. To increase power for variants discovered via genome sequencing, we imputed sequence-based genotypes into 44,414 additional individuals of European origin (11,645 T2D cases and 32,769 controls; Methods) from 13 studies (Supplementary Table 3). We estimated power in the combined sequence plus imputed data, adjusting for imputation

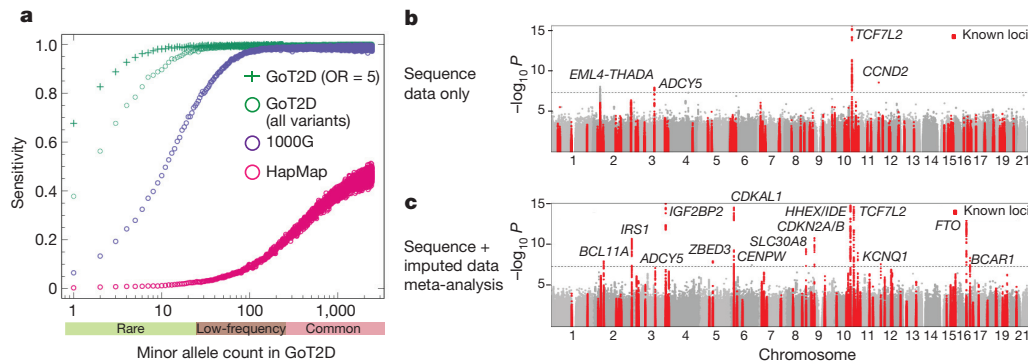


Figure 1 | Ascertainment of variants and single-variant results.

a, Sensitivity of low-coverage genome sequence data to detect SNVs in the deep exome sequence data, relative to other variant catalogues. Points represent results for a specific minor allele count. All results assume odds ratio (OR) = 1 for all variants, unless stated otherwise.

quality, to be $\geq 80\%$ for variants with $MAF \geq 5\%$ and $OR \geq 1.23$, or $MAF \geq 0.5\%$ and $OR \geq 1.92$ (Extended Data Fig. 2). A meta-analysis combining results for the sequence and imputed data identified 674 variants across 14 loci associated with T2D at genome-wide significance (Fig. 1c). We observed a previously undescribed association with a common variant near *CENPW* (rs11759026, $MAF = 23.2\%$, $P_{meta} = 3.5 \times 10^{-8}$; Fig. 1c) and replicated this association in an additional 14,201 cases and 100,964 controls from the DIAGRAM consortium ($P = 2.5 \times 10^{-4}$; $P_{combined} = 1.1 \times 10^{-11}$; Methods). The *EML4* signal detected in the sequence data was not replicated in the imputed data ($P = 0.59$; $P_{meta} = 0.26$; Fig. 1c).

To test for additional association signals, we performed conditional analysis at loci previously associated with risk of T2D (Methods). We identified two previously unreported association signals, both involving low-frequency variants, at a corrected significance threshold ($\alpha < 1.8 \times 10^{-6}$; Methods): one at the *IRS1* locus (rs78124264, $MAF = 2.2\%$, $P_{conditional} = 2.5 \times 10^{-7}$) and one upstream of *PPARG* (rs79856023, $MAF = 2.2\%$, $P_{conditional} = 9.2 \times 10^{-7}$) (Extended Data Table 3). The *PPARG* signal overlaps regulatory elements in hASC pre-adipose and HepG2 cells, consistent with evidence that altered adipose regulation drives the primary *PPARG* signal¹⁴.

Analysis of coding variation

The T2D-GENES consortium adopted a complementary strategy, focused on variants in protein-coding sequence, and seeking to improve power to detect rare-variant association by exploiting the more robust functional annotation of coding variation and the potential to aggregate multiple alleles of presumed similar impact in the same gene^{12,15}. We combined exome sequence data from 10,437 T2D cases and controls of diverse ancestry generated by T2D-GENES with the equivalent data from GoT2D. This created a joint data set (after all quality control) comprising 12,940 individuals (6,504 cases and 6,436 controls) drawn from five ancestry groups: 4,541 of European origin, and around 2,000 (range: 1,943–2,217) each of South Asian, East Asian, Hispanic and African American origin (Extended Data Fig. 1, Extended Data Table 1 and Supplementary Table 4). Mean coverage was $82\times$ across the coding sequence of 18,281 genes, identifying 3.04 million variants (1.19 million protein-altering; Supplementary Figs 5, 6). Each diploid genome carried a mean of 9,243 (range: 8,423–11,487) synonymous, 7,636 (6,935–9,271) missense, and 250 (183–358) protein-truncating alleles (Supplementary Table 7).

We tested for T2D association within the five ancestry groups, assuming an additive genetic model, using mixed-model approaches that account for population structure and relatedness¹⁶, and combined ancestry-specific results by trans-ethnic meta-analysis (Methods). We estimate $\geq 80\%$ power to detect (at genome-wide significance) T2D risk variants with $MAF \geq 5\%$ and $OR \geq 1.36$, or

b, c, Manhattan plots of single-variant association analyses for: sequence data alone (**b**, 1,326 cases and 1,331 controls) and meta-analysis of sequence and imputed data (**c**, total of 14,297 cases and 32,774 controls). 1000G, the 1000 Genomes Project data.

$MAF \geq 0.5\%$ and $OR \geq 2.29$ (Methods and Extended Data Fig. 2). Only one variant reached genome-wide significance (*PAX4* Arg192His, rs2233580, $P = 9.3 \times 10^{-9}$) (Table 1, Extended Data Figs 3, 4 and Supplementary Fig. 8). This association was exclusive to East Asian individuals, in whom the 192His allele is common ($MAF \approx 10\%$) with a substantial effect size (allelic $OR = 1.79$ (1.47–2.19)); 192His is virtually absent in individuals from other ancestries ($MAF = 0.014\%$). The rs2233580 association was replicated in independent East Asian case-control data ($n = 3,301$; $P = 5.9 \times 10^{-7}$; Supplementary Table 9) and was distinct ($r^2 < 0.05$) from previously reported GWAS SNVs at the *GCC1–PAX4* locus^{6,8}. *PAX4* encodes a transcription factor involved in islet differentiation and function¹⁷ (Supplementary Table 10), and *PAX4* variants have been implicated in early-onset monogenic diabetes¹⁸. However, in East Asian cases, 192His was not associated with age of diabetes diagnosis ($P = 0.64$), indicating that this variant influences risk of type 2 diabetes rather than early-onset monogenic diabetes (Supplementary Table 9).

To increase power to detect the association of rare variants that cluster in individual genes, we deployed gene-level variant aggregation tests¹⁵ across the exome sequence data (Methods and Supplementary Table 11). We observed no deviation from the null distribution of association statistics, and no single gene reached exome-wide significance ($\alpha = 2.5 \times 10^{-6}$; Methods and Supplementary Figs 12, 13). When we focused on 634 genes that mapped to known GWAS regions, only *FES* exceeded a reduced significance threshold of $\alpha = 7.9 \times 10^{-5}$ ($P_{SouthAsian} = 7.2 \times 10^{-6}$; $P_{multiethnic} = 1.9 \times 10^{-5}$) (Methods and Supplementary Fig. 14). This aggregate signal was driven entirely by the South Asian-specific Pro536Ser variant ($MAF = 0.9\%$, $OR = 6.7$ (2.6–17.3), $P = 7.5 \times 10^{-6}$), indicating that *FES* is likely to be the effector gene at the *PRC1* GWAS locus⁴.

To increase power to detect coding variant associations (Extended Data Fig. 2), we contributed early T2D-GENES exome data to the design of an Illumina exome array⁹, and then collected genotypes from an additional 28,305 T2D cases and 51,549 controls of European ancestry from 13 studies (Extended Data Fig. 1, Extended Data Table 1 and Supplementary Table 15). Of 27,904 protein-altering variants with $MAF > 0.5\%$ detected in exome sequence data from 4,541 European individuals, variation at 81.6% was captured on the array (Supplementary Fig. 16).

Association analysis in the combined sequence and array data from more than 90,000 individuals identified 18 coding variants (17 non-synonymous) at 13 loci that exceeded genome-wide significance ($\alpha = 5 \times 10^{-8}$; Table 1 and Extended Data Figs 3, 4). All of these were common ($MAF > 5\%$) and all but one mapped within established common-variant GWAS regions^{2,3}. The exception, which we replicated in the INTERACT study¹⁹ ($n = 9,292$; $P_{INTERACT} = 2.4 \times 10^{-4}$; $P_{meta} = 2.2 \times 10^{-11}$), involved a common haplotype of four strongly

Table 1 | Nonsynonymous coding variants achieving genome-wide significance

Locus	Gene	Variant	RAF range	Eur MAF	Alleles	Exomes (<i>n</i> = 12,940)		Exome-chip (<i>n</i> = 79,854)		Combined (<i>n</i> = 92,794)	
						<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)
Established common causal coding variant signals											
<i>GCKR</i>	<i>GCKR</i>	rs1260326 Pro446Leu	0.49–0.86	0.37	C, T	0.075	1.05 (0.99–1.11)	4.8 × 10 ^{−9}	1.07 (1.04–1.11)	1.2 × 10 ^{−9}	1.07 (1.04–1.10)
<i>PPARG</i>	<i>PPARG</i>	rs1801282 Pro12Ala	0.86–0.99	0.14	C, G	0.0030	1.16 (1.06–1.27)	1.8 × 10 ^{−7}	1.10 (1.06–1.14)	4.2 × 10 ^{−8}	1.11 (1.07–1.15)
<i>PAM/ PPIP5K2</i>	<i>PAM</i>	rs35658696 Asp563Gly	0.00–0.05	0.054	G, A	0.00045	1.36 (1.14–1.63)	1.7 × 10 ^{−7}	1.15 (1.08–1.23)	5.7 × 10 ^{−10}	1.17 (1.11–1.24)
	<i>PPIP5K2</i>	rs36046591 Ser1207Gly	0.00–0.05	0.054	G, A	0.0099	1.34 (1.12–1.61)	1.0 × 10 ^{−6}	1.17 (1.10–1.25)	3.3 × 10 ^{−8}	1.19 (1.12–1.26)
<i>SLC30A8</i>	<i>SLC30A8</i>	rs13266634 Arg325Trp	0.58–0.91	0.33	C, T	2.9 × 10 ^{−6}	1.15 (1.09–1.22)	2.7 × 10 ^{−18}	1.14 (1.11–1.17)	4.8 × 10 ^{−23}	1.14 (1.11–1.17)
<i>KCNJ11/ ABCC8</i>	<i>KCNJ11</i>	rs5215 Val337Ile	0.08–0.40	0.40	C, T	0.11	1.07 (1.01–1.13)	3.4 × 10 ^{−9}	1.07 (1.04–1.11)	1.3 × 10 ^{−9}	1.07 (1.05–1.10)
		rs5219 Lys23Glu	0.06–0.40	0.40	T, C	0.056	1.08 (1.02–1.14)	5.1 × 10 ^{−9}	1.07 (1.04–1.11)	9.0 × 10 ^{−10}	1.07 (1.05–1.10)
	<i>ABCC8</i>	rs757110 Ala1369Ser	0.06–0.40	0.40	C, A	0.20	1.06 (1.00–1.12)	2.3 × 10 ^{−8}	1.07 (1.04–1.11)	1.7 × 10 ^{−8}	1.07 (1.04–1.10)
Other coding variant associations within established common variant GWAS regions											
<i>THADA</i>	<i>THADA</i>	rs35720761 Cys1605Tyr	0.85–1.00	0.10	C, T	0.0021	1.12 (1.01–1.23)	3.5 × 10 ^{−8}	1.11 (1.07–1.16)	3.3 × 10 ^{−10}	1.12 (1.07–1.16)
<i>COBLL1</i>	<i>COBLL1</i>	rs7607980 Asn939Asp	0.84–1.00	0.12	T, C	1.4 × 10 ^{−5}	1.21 (1.11–1.33)	4.7 × 10 ^{−11}	1.14 (1.10–1.19)	8.3 × 10 ^{−15}	1.15 (1.11–1.19)
<i>WFS1</i>	<i>WFS1</i>	rs1801212 Val333Ile	0.70–1.00	0.30	A, G	0.0026	1.14 (1.06–1.23)	9.3 × 10 ^{−12}	1.08 (1.04–1.12)	9.0 × 10 ^{−14}	1.09 (1.06–1.12)
		rs1801214 Asn500Asn	0.59–0.96	0.41	T, C	0.0019	1.08 (1.02–1.15)	2.0 × 10 ^{−12}	1.08 (1.05–1.11)	1.5 × 10 ^{−14}	1.08 (1.05–1.11)
		rs734312 Arg611His	0.11–0.85	0.47	A, G	0.12	1.05 (0.99–1.11)	1.3 × 10 ^{−10}	1.07 (1.03–1.10)	6.9 × 10 ^{−11}	1.06 (1.04–1.09)
<i>RREB1</i>	<i>RREB1</i>	rs9379084 Asp1171Asn	0.87–0.98	0.11	G, A	2.2 × 10 ^{−5}	1.19 (1.09–1.30)	1.1 × 10 ^{−5}	1.12 (1.06–1.17)	4.0 × 10 ^{−9}	1.13 (1.09–1.18)
<i>PAX4</i>	<i>PAX4</i>	rs2233580 Arg192His	0.00–0.10	0.00	T, C	9.3 × 10 ^{−9}	1.79 (1.47–2.19)	NA	NA	9.3 × 10 ^{−9}	1.79 (1.47–2.19)
<i>GPSM1</i> *	<i>GPSM1</i> *	rs60980157 Ser391Leu	0.26	0.26	C, T	NA	NA	1.7 × 10 ^{−9}	1.09 (1.06–1.12)	1.7 × 10 ^{−9}	1.09 (1.06–1.12)
<i>CILP2</i>	<i>TM6SF2</i>	rs58542926 Glu167Lys	0.03–0.10	0.082	T, C	0.00015	1.22 (1.10–1.36)	1.9 × 10 ^{−7}	1.13 (1.08–1.18)	3.2 × 10 ^{−10}	1.14 (1.10–1.19)
Coding variant associations outside established common variant GWAS regions											
<i>MTMR3/b ASCC2</i>	<i>MTMR3</i>	rs41278853 Asn960Ser	0.92–1.00	0.083	A, G	9.2 × 10 ^{−5}	1.26 (1.12–1.42)	3.2 × 10 ^{−6}	1.12 (1.07–1.17)	5.6 × 10 ^{−9}	1.14 (1.09–1.19)
	<i>ASCC2</i>	rs11549795 Val123Ile	0.92–1.00	0.083	C, T	0.00040	1.23 (1.10–1.38)	2.0 × 10 ^{−5}	1.11 (1.06–1.16)	1.0 × 10 ^{−7}	1.13 (1.08–1.18)
		rs28265 Asp407His	0.92–1.00	0.083	C, G	0.00050	1.21 (1.08–1.36)	1.9 × 10 ^{−5}	1.11 (1.06–1.16)	1.1 × 10 ^{−7}	1.12 (1.08–1.17)
		rs36571 Pro423Ser	0.92–1.00	0.083	G, A	0.0023	1.23 (1.08–1.40)	2.0 × 10 ^{−5}	1.11 (1.06–1.16)	3.0 × 10 ^{−7}	1.12 (1.08–1.17)

These loci were identified through single-variant analyses of exome sequence data from 6,504 cases and 6,436 controls and exome arrays from 28,305 cases and 51,549 controls. RAF, risk allele frequency; Eur MAF, minor allele frequency in Europeans; OR, odds ratio; CI, confidence interval; *n*, total number of individuals analysed. Genome-wide significance defined as $P < 5 \times 10^{-8}$.

**GPSM1* variant failed quality control in exome sequence: association *P* values derive only from exome-array analysis. The synonymous variant Thr515Thr (rs55834942) in *HNF1A* also reached genome-wide significance ($P = 1.0 \times 10^{-8}$) in the combined analysis. Three variants at *ASCC2* did not achieve genome-wide significance themselves, but are included because they fall into a region with an other significant association signal. Alleles are aligned to the forward strand of NCBI Build 37 and represented as risk and other alleles.

correlated coding variants in *MTMR3* and *ASCC2* (Table 1). Of these variants, *MTMR3* Asn960Ser (MAF = 8.3%) had the strongest residual association signal on conditional analysis, implicating *MTMR3*, which encodes a phosphatidylinositol phosphatase²⁰, as the probable effector transcript at this locus (Extended Data Table 3, Extended Data Figs 3, 4, Supplementary Table 10 and Supplementary Fig. 17).

The remaining coding variant signals provided an opportunity to highlight causal alleles and effector transcripts for known GWAS signals. For five loci (*SLC30A8*, *GCKR*, *PPARG*, *KCNJ11*–*ABCC8*, and *PAM*), the coding variants identified had previously been nominated as causal for their respective GWAS signals^{2,7,13}. For the other seven loci, GWAS meta-analyses had previously highlighted a lead variant

in non-coding sequence^{2,5,6}. We (re)evaluated these relationships with conditional and credible set analyses and found that, at most, the evidence supported a direct causal role for the coding variants concerned (Extended Data Table 3, Extended Data Figs 3, 4, Supplementary Table 10 and Supplementary Fig. 17).

For example, at the *CILP2* locus², previous GWAS had identified the non-coding variant rs10401969 as the lead SNV. However, direct genotyping of *TM6SF2* Glu167Lys on the exome array revealed complete linkage disequilibrium with rs10401969, and reciprocal signal extinction in conditional analyses (Extended Data Table 3 and Extended Data Figs 3, 4). In previous GWAS, the association at Glu167Lys had been obscured by incomplete genotyping and poor imputation

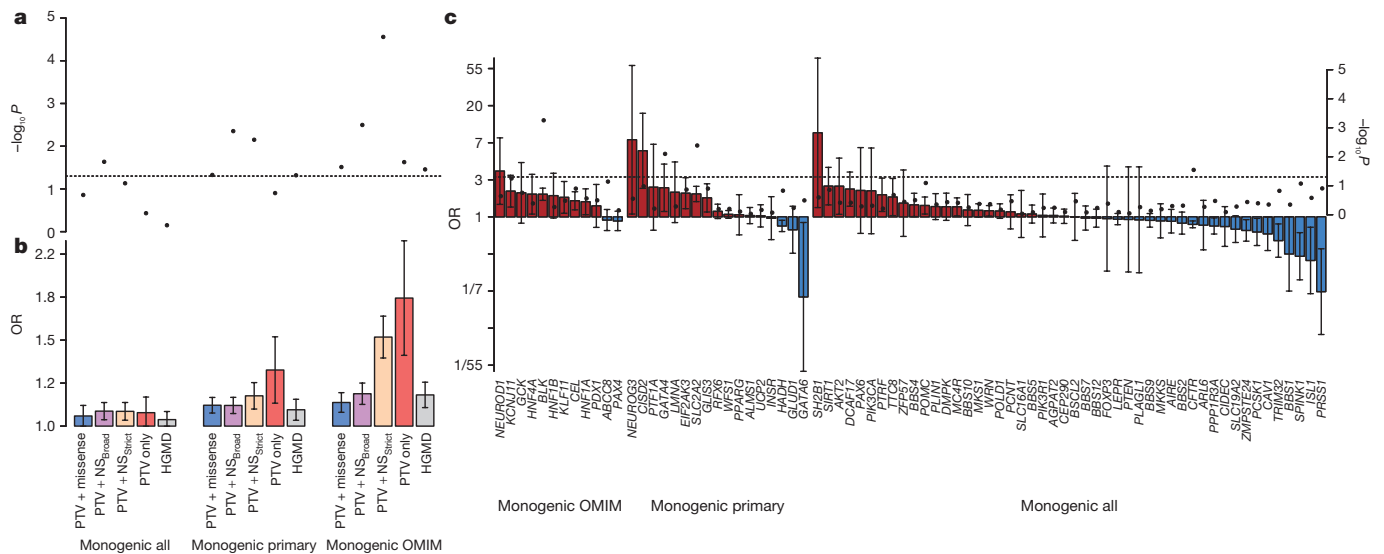


Figure 2 | Association between T2D and variants in genes for Mendelian forms of diabetes. **a**, P values (plotted as $-\log_{10} P$) of aggregate association for variants from 6,504 T2D cases and 6,436 controls in three sets of Mendelian diabetes genes, for five variant ‘masks’ (see Methods). Dotted line, $P = 0.05$. **b**, Estimated T2D odds ratio (OR) for carriers of

variants in each gene-set and mask. **c**, Estimated ORs (bars, left axis) and P values (dots, right axis) for carriers of variants in the PTV + NS_{strict} mask for each gene. Red, OR > 1; blue, OR < 1; dotted line, $P = 0.05$. Error bars represent s.e.

(Supplementary Table 18). The *TM6SF2* Lys167 allele has been shown to underlie predisposition to hepatic steatosis²¹, and was associated with fasting hyperinsulinaemia ($P = 1.0 \times 10^{-4}$) in 30,824 non-diabetic controls from the present study. This combination of genetic and functional data, consistent with known mechanistic links between insulin resistance, T2D, and fatty liver disease²², implicates *TM6SF2* Glu167Lys as the most likely T2D risk variant at this locus.

By contrast, the association at *RREB1* Asp1171Asn represented a novel signal, conditionally independent of the adjacent common-variant GWAS signal. This association, together with that involving a second associated coding variant, Ser1554Tyr, which has a marked association with fasting glucose levels ($P = 2.7 \times 10^{-9}$ in 38,338 non-diabetic subjects from the present study; Supplementary Table 19), establishes *RREB1* (ref. 23) as the probable effector gene at the *SSR1* locus.

Given the concentration of coding-variant associations within established GWAS loci, we sought to nominate additional single-variant signals in 634 genes that mapped to established T2D GWAS regions using a Bonferroni-corrected $\alpha = 1.6 \times 10^{-5}$ (Methods, Supplementary Fig. 14 and Supplementary Table 20). At *HNF4A*, we confirmed a T2D association at Thr139Ile (European MAF range 0.7–3.8%, OR = 1.15 (1.08–1.22), $P = 2.9 \times 10^{-6}$)¹⁰ that was distinct from both the common non-coding lead GWAS SNV^{2,3,5} and multiple rare *HNF4A* variants implicated in monogenic diabetes²⁴. Additional coding variant associations in *TSPAN8* and *THADA* highlighted these two genes as probable effector transcripts in their respective GWAS regions (Supplementary Tables 10, 21).

Rare alleles in Mendelian genes

We extended gene-based tests for rare-variant associations to gene-sets implicated in monogenic or syndromic diabetes or in altered glucose metabolism²⁴. Across 81 genes harbouring rare alleles that are causal for monogenic or syndromic diabetes or related glycaemic traits (‘Monogenic all’; Supplementary Table 22), the only variant or gene association achieving genome-wide significance involved the previously mentioned *PAX4* Arg192His. However, across the entire gene-set, we observed a weak aggregate association with T2D risk ($P = 0.023$; Fig. 2a). The association was considerably stronger in two subsets of genes that have been more directly implicated in monogenic and syndromic diabetes: a manually curated set of 28 genes for which diabetes was the primary phenotype (‘Monogenic primary’) and a partially overlapping

set of 13 genes reported in OMIM (<http://www.omim.org>) as causal for maturity onset diabetes of the young (MODY) or neonatal diabetes (‘Monogenic OMIM’; Supplementary Table 22).

The ‘Monogenic OMIM’ gene-set had a statistically robust signal of association ($P = 2.8 \times 10^{-5}$, OR = 1.51 (1.25–1.83)) driven by the allelic burden of MAF < 1% alleles. Effect size estimates tracked with increasing stringency of variant annotation and gene-set definition, consistent with progressive enrichment for functional over neutral alleles (Fig. 2b). This signal does not reflect inclusion among T2D cases of individuals who, in reality, had monogenic diabetes: the association was not concentrated among genes most frequently responsible for monogenic diabetes²⁴ (Fig. 2c), and age of diabetes diagnosis was no lower in variant carriers than non-carriers (Supplementary Fig. 23). The association signal remained after all alleles listed as ‘disease-causing’ within the Human Genetic Mutation Database were excluded ($P = 2.9 \times 10^{-4}$, OR = 1.50 (1.21–1.86)).

These analyses point to widespread enrichment for T2D association among rare coding alleles in genes that are causal for monogenic diabetes. In these genes, alleles of penetrance sufficient to drive familial segregation of early-onset diabetes coexist alongside those with more modest effects predisposing carriers to later-onset T2D. No other compelling signals of rare-variant enrichment were detected using gene-set enrichment or protein–protein interaction analysis in other pre-defined gene-sets (Supplementary Figs 24, 26 and Supplementary Table 25).

No evidence for synthetic association

In 2010, Goldstein and colleagues proposed that common-variant GWAS signals may be the consequence of low-frequency and rare variants that by chance cluster on common haplotypes²⁵. Although this hypothesis has been debated^{26,27} and assessed indirectly^{3,28}, we used the near-complete ascertainment of genetic variation in 2,657 genome-sequenced individuals to directly test the importance of ‘synthetic’ associations²⁹. We focused on the ten T2D GWAS loci at which our sample provided the strongest statistical evidence for association ($P < 0.001$), implementing a conditional analysis procedure to assess whether combinations of SNVs within a 5-Mb window could explain the common-variant signal (Extended Data Table 4 and Methods).

We first focused on missense variants, finding that none of the ten signals could be explained by low-frequency and rare variants within 2.5 Mb of the common index SNV (Extended Data Fig. 5). For example,

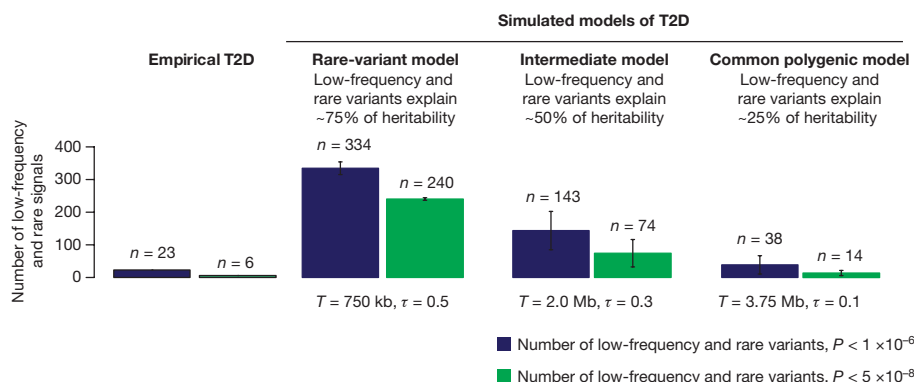


Figure 3 | Empirical T2D association results compared to results under different simulated disease models. Observed number of rare and low-frequency (MAF < 5%) genetic association signals for T2D detected genome-wide after imputation compared to the numbers seen under three simulated disease models for T2D which were plausible given results (T2D recurrence risks, GWAS, linkage) before large-scale sequencing.

at the *IRS1* locus, including the five observed missense *IRS1* alleles in the model did not meaningfully diminish the index SNV association ($P_{\text{unconditional}} = 2.8 \times 10^{-6}$; $P_{\text{conditional}} = 4.3 \times 10^{-6}$). With 99.7% ascertainment of low-frequency coding variants (Methods), these results rule out synthetic associations produced by missense variants at these ten loci.

We expanded the search to include all low-frequency and rare variants, non-coding and coding, within 2.5 Mb of index SNVs. At no locus was a single low-frequency or rare variant sufficient to explain the GWAS signal (Extended Data Fig. 5). At eight of the ten loci, ten or more low-frequency and rare variants were needed to reverse the direction of effect at the common index SNV; at *TCF7L2*, even 50 were insufficient (Extended Data Fig. 5). We note that the statistical procedure we developed and deployed is biased *in favour* of the synthetic association hypothesis, since it is highly prone to over-fitting. Nonetheless, at 8 of the 10 loci the data were indistinguishable from a null model of no synthetic association (Extended Data Table 4 and Supplementary Fig. 27).

Nominating candidate functional alleles

Using the GoT2D whole-genome sequence data, we constructed 99% 'credible sets' for each T2D GWAS locus on the assumption that there is one causal variant per locus³⁰ (Methods). Across 78 published autosomal loci at which the reported index SNV had MAF > 1%, 99% credible set sizes ranged from 2 (*CDKN2A/B*) to around 1,000 (*POU5F1*) variants; at 71 loci, the credible set contained more than 10 variants (Extended Data Fig. 5 and Supplementary Table 28). The GoT2D data set provides near-complete ascertainment of common and low-frequency variants, which support more comprehensive credible set analysis than studies based on genotyping or imputation alone^{3,31}; of the credible set variants identified from whole-genome sequence data, about 60% are absent from HapMap and ~5% from 1000G Phase 1 (Extended Data Fig. 5).

Genomic maps of chromatin state or transcription factor binding^{32–35} have been used to prioritize causal variants within credible sets^{36,37}. We jointly modelled genetic association and genomic annotation data at T2D GWAS loci using fgwas³⁸. Consistent with previous reports^{34,35}, associated variants were enriched in coding exons, transcription factor binding sites, and enhancers active in pancreatic islets and adipose tissue (Extended Data Fig. 6). Overall, including the functional annotation data reduced the credible set size by 35%. At several loci, access to complete sequence data prioritized variants that overlapped with relevant regulatory annotations and had been previously overlooked. For example, at the *CCND2* locus, three variants not present in HapMap Phase 2 have a combined probability of 90.0% of explaining the common-variant signal² (Extended Data Fig. 6); one of these (rs3217801) is a 2-base pair (bp) indel overlapping an islet enhancer element.

Simulated models were defined by two parameters: disease target size T and degree of coupling τ between the causal effects of variants and the selective pressure against them⁴⁰. Simulated data were generated to match GoT2D imputation quality as a function of MAF (see Methods). Error bars represent s.e. observed across simulation replicates (bar value shows the mean).

Modelling disease architecture

To evaluate the overall contribution of low-frequency coding variation to T2D risk, we estimated the proportion of variance in T2D liability attributable to each such variant³⁹ (Methods and Extended Data Fig. 7). We focused on exome array data to maximize sample size, and on variants with MAF > 0.1%; the sensitivity of variant ascertainment and accuracy of OR estimation decline below this threshold. Among the 31,701 variants on the exome array with $0.1\% < \text{MAF} < 5\%$, there was a progressive increase in the maximum OR estimates with decreasing frequency. However, the liability variance explained for these variants rarely exceeded 0.05%, limiting the power to detect association in the sample size available (Extended Data Fig. 7). We estimated (Methods) that the liability variance that was collectively attributable to coding variants in the $0.1\% < \text{MAF} < 5\%$ range was 2.9%, compared to 6.3% for common variants.

Finally, we compared our whole-genome T2D association results with predictions from population genetic simulations⁴⁰ under twelve models that varied widely with respect to the proportion of heritability explained by common, low-frequency, and rare variants. We mirrored the GoT2D study design (with imputation) and performed in parallel the same association analysis on empirical and simulated data, focusing on variants with MAF > 0.1% and allowing for power loss due to imperfect imputation (Methods).

Figure 3 displays results for three representative models: a 'purifying selection' model in which low-frequency and rare variants explain approximately 75% of T2D heritability; an intermediate model in which both common and lower-frequency variants contribute substantially; and a 'neutral' model in which common variants explain about 75% of T2D heritability. The predictions of the first two models differ markedly from the empirical data with respect to the numbers of low-frequency and rare risk variants that are associated with T2D. Specifically, these two models predict a larger number and greater effect size of low-frequency variants should be found in our whole-genome sequencing study as compared to those observed in the empirical data. By contrast, the empirical data are consistent with predictions under the 'neutral' common-variant model.

The century-old Mendelian–biometrician debate pitted those who attributed trait variation to rare variants of large effect against those who argued that trait variation was largely due to many common variants of small effect. The debate today is about whether the 'missing heritability' after GWAS is due largely to individually rare, highly penetrant variants⁴¹ or to a large universe of common alleles with modest effects⁴². The results are of more than academic interest, as genetic architecture plays out powerfully in relation to the power of genetic diagnosis and the application of precision medicine.

Our data and analysis indicate that, for T2D, nearly all common-variant associations detectable by whole genome sequencing were previously found by GWAS based on genotyping arrays and imputation; concerns about incomplete coverage due to 'holes' in HapMap¹¹ coverage were, we show, unfounded. Of more lasting interest, the combination of genome and exome sequencing in large samples provides limited evidence of a role for lower-frequency variants—coding or genome-wide—in T2D predisposition. Of course, rare risk alleles have long been known to contribute in families with early-onset forms of diabetes, and sequencing of Mendelian and GWAS genes has identified rare variants that influence disease risk^{43,44}. Sequencing of T2D cases in much larger samples will undoubtedly uncover additional low-frequency and rare variants that provide biological and potentially clinical value. Nonetheless, our empirical and simulated data suggest that these lower-frequency variants contribute much less to T2D heritability than do common variants. Moreover, the frequency spectrum of variant association signals is consistent with a model in which limited selective pressure distributes most of the genetic variance influencing T2D risk among common alleles⁴⁰, in line with the frequency distribution of inter-individual sequence variation. Similar large-scale sequencing-based exploration of other complex traits will be required to determine the extent to which the genetic architecture of T2D is representative of other late-onset diseases.

Our results further strengthen the case for sequencing of diverse samples; the population-enriched T2D risk variant in *PAX4* dovetails with similar findings involving *SLC16A11* (ref. 45) in East Asian and Native American populations and *TBC1D4* (ref. 46) in Greenland Inuits. Studies involving populations that have been subject to bottlenecks or extreme selective pressures^{43,46,47} may be particularly fruitful.

Understanding the inherited basis of T2D will require much further progress in identifying the mechanisms whereby common, mostly non-coding, variants influence disease risk. The combination of global epigenetic measurements, genome editing⁴⁸, and high-throughput functional assays⁴⁹ make it increasingly practical to characterize large numbers of non-coding variants and the processes they affect. Genome sequencing in much larger numbers of individuals than included in the current study are needed and will no doubt provide foundational information to guide such experimentation and connect the results to human population variation, physiology, and disease. Integration of biological insights gleaned from common and rare variant associations with T2D into a unified picture of disease pathophysiology will be required to fully understand the basis of this common but challenging disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 May 2015; accepted 12 June 2016.

Published online 11 July 2016.

- Willemsen, G. *et al.* The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the discordant twin (DISCOTWIN) consortium. *Twin Res. Hum. Genet.* **18**, 762–771 (2015).
- Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
- Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
- Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
- Kooner, J. S. *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* **43**, 984–989 (2011).
- Cho, Y. S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat. Genet.* **44**, 67–72 (2011).
- Steinthorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46**, 294–298 (2014).
- Ma, R. C. *et al.* Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near *PAX4*. *Diabetologia* **56**, 1291–1305 (2013).
- Huyghe, J. R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45**, 197–201 (2013).
- Gaulton, K. J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
- Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Lohmueller, K. E. *et al.* Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. *Am. J. Hum. Genet.* **93**, 1072–1086 (2013).
- Albrechtsen, A. *et al.* Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* **56**, 298–310 (2013).
- Claussnitzer, M. *et al.* Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* **156**, 343–358 (2014).
- Lee, S., Teslovich, T. M., Boehnke, M. & Lin, X. General framework for meta-analysis of rare variants in sequencing association studies. *Am. J. Hum. Genet.* **93**, 42–53 (2013).
- Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
- Collombat, P. *et al.* Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev.* **17**, 2591–2603 (2003).
- Kooptiwut, S. *et al.* Defective *PAX4* R192H transcriptional repressor activities associated with maturity onset diabetes of the young and early onset-age of type 2 diabetes. *J. Diabetes Complications* **26**, 343–347 (2012).
- Langenberg, C. *et al.* Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54**, 2272–2282 (2011).
- Oppelt, A. *et al.* Production of phosphatidylinositol 5-phosphate via PIKfyve and MTMR3 regulates cell migration. *EMBO Rep.* **14**, 57–64 (2013).
- Kozlitina, J. *et al.* Exome-wide association study identifies a *TM6SF2* variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46**, 352–356 (2014).
- Mahdessian, H. *et al.* *TM6SF2* is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc. Natl Acad. Sci. USA* **111**, 8913–8918 (2014).
- Thiagalingam, A., Lengauer, C., Baylin, S. B. & Nelkin, B. D. RREB1, a ras responsive element binding protein, maps to human chromosome 6p25. *Genomics* **45**, 630–632 (1997).
- Murphy, R., Ellard, S. & Hattersley, A. T. Clinical implications of a molecular genetic classification of monogenic β -cell diabetes. *Nat. Clin. Pract. Endocrinol. Metab.* **4**, 200–213 (2008).
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
- Anderson, C. A., Soranzo, N., Zeggini, E. & Barrett, J. C. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol.* **9**, e1000580 (2011).
- Wray, N. R., Purcell, S. M. & Visscher, P. M. Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol.* **9**, e1000579 (2011).
- Sim, X. *et al.* Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet.* **7**, e1001363 (2011).
- Goldstein, D. B. The importance of synthetic associations will only be resolved empirically. *PLoS Biol.* **9**, e1001008 (2011).
- Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
- Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Mikkelsen, T. S. *et al.* Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156–169 (2010).
- Parker, S. C. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
- Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
- Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).
- Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
- Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
- Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
- Agarwala, V., Flannick, J. & Sunyaev, S. GoT2D Consortium & Altshuler, D. Evaluating empirical bounds on complex disease genetic architecture. *Nat. Genet.* **45**, 1418–1427 (2013).

41. McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210–217 (2010).
42. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
43. Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).
44. Bonnefond, A. *et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* **44**, 297–301 (2012).
45. Sigma Type 2 Diabetes Consortium *et al.* Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97–101 (2014).
46. Moltke, I. *et al.* A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
47. Sigma Type 2 Diabetes Consortium *et al.* Association of a low-frequency variant in *HNFI1A* with type 2 diabetes in a Latino population. *JAMA* **311**, 2305–2314 (2014).
48. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
49. Majithia, A. R. *et al.* Rare variants in *PPARG* with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proc. Natl Acad. Sci. USA* **111**, 13127–13132 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements Grant support and acknowledgments are listed in the Supplementary Information.

Author Contributions Author contributions are described in the Supplementary Information.

Author Information Whole-genome sequence data from the GoT2D project are available by application to the European Genome-Phenome Archive (<https://www.ebi.ac.uk/ega/home>) under accession number EGAS00001001459 and from dbGAP (<http://www.ncbi.nlm.nih.gov/gap>) under accession number phs000840.v1.p1. Whole-exome sequence data from the T2D-GENES project are available from the European Genome-Phenome Archive (<https://www.ebi.ac.uk/ega/home>) under accession number EGAS00001001460 and from dbGAP (<http://www.ncbi.nlm.nih.gov/gap>) under accession numbers phs000847.v1.p1, phs001093.v1.p1, phs001095.v1.p1, phs001096.v1.p1, phs001097.v1.p1, phs001098.v1.p1, phs001099.v1.p1, phs001100.v1.p1 and phs001102.v1.p1. Summary-level data from the exome array component of this project (and from the exome and genome sequences) can be freely accessed at the Accelerating Medicines Partnership T2D portal (<http://www.type2diabetesgenetics.org>), and similar data from the GoT2D-imputed data at <http://www.diagram-consortium.org>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.I.M. (mark.mccarthy@dr1.ox.ac.uk) or M.B. (boehnke@umich.edu).

- Christian Fuchsberger^{1,2,3*}, Jason Flannick^{4,5*}, Tanya M. Teslovich^{1*}, Anubha Mahajan^{6*}, Vineeta Agarwala^{4,7*}, Kyle J. Gaulton^{6*}, Clement Ma^{1†}, Pierre Fontanillas^{4†}, Loukas Moutsianas^{6†}, Davis J. McCarthy^{6,8}, Manuel A. Rivas⁶, John R. B. Perry^{6,9,10,11}, Xueling Sim^{1†}, Thomas W. Blackwell¹, Neil R. Robertson^{6,12}, N. William Rayner^{6,12,13}, Pablo Cingolani^{14,15}, Adam E. Locke^{1†}, Juan Fernandez-Tajes⁶, Heather M. Highland^{16†}, Josee Dupuis^{17,18}, Peter S. Chines¹⁹, Cecilia M. Lindgren^{4,6†}, Christopher Hartl⁴, Anne U. Jackson¹, Han Chen^{17,20}, Jeroen R. Huyghe^{1†}, Martijn van de Bunt^{6,12}, Richard D. Pearson⁶, Ashish Kumar^{6,21}, Martina Müller-Nurasyid^{22,23,24,25}, Niels Garup²⁶, Heather M. Stringham¹, Eric R. Gamazon^{27†}, Jaehoon Lee²⁸, Yuhui Chen⁶, Robert A. Scott¹⁰, Jennifer E. Below²⁹, Peng Chen³⁰, Jinyan Huang^{31†}, Min Jin Go³², Michael L. Stitzel³³, Dorota Pasko⁹, Stephen C. J. Parker³⁴, Tibor V. Varga³⁵, Todd Green⁴, Nicola L. Beer¹², Aaron G. Day-Williams^{13†}, Teresa Ferreira⁶, Tasha Fingerlin³⁶, Momoko Horikoshi^{6,12}, Cheng Hu³⁷, Iksoo Huh²⁸, Mohammad Kamran Ikram^{38,39,40}, Bong-Jo Kim³², Yongkang Kim²⁸, Young Jin Kim³², Min-Seok Kwon⁴¹, Juyoung Lee³², Selyeong Lee²⁸, Keng-Han Lin¹, Taylor J. Maxwell^{29†}, Yoshihiko Nagai^{15,42,43}, Xu Wang³⁰, Ryan P. Welch¹, Joon Yoon^{4†}, Weihua Zhang^{44,45}, Nir Barzilai⁴⁶, Benjamin F. Voight^{47,48}, Bok-Ghee Han³², Christopher P. Jenkinson^{49,50}, Teemu Kuulasmaa⁵¹, Johanna Kuusisto^{51,52}, Alisa Manning⁴, Maggie C. Y. Ng⁵³, Nicholette D. Palmer^{53,54,55}, Beverley Balkau⁵⁶, Alena Stančáková⁵¹, Hanna E. Abboud^{49†}, Heiner Boeing⁵⁷, Vilmantas Giedraitis⁵⁸, Dorairaj Prabhakaran⁵⁹, Omri Gottesman⁶⁰, James Scott⁶¹, Jason Carey⁴, Phoenix Kwan¹, George Grant⁴, Joshua D. Smith⁶², Benjamin M. Neale^{63,64}, Shaun Purcell^{4,64,65}, Adam S. Butterworth⁶⁶, Joanna M. M. Howson⁶⁶, Heung Man Lee⁶⁷, Yingchang Lu⁶⁰, Soo-Heon Kwak⁶⁸, Wei Zhao⁶⁹, John Danesh^{13,66,70}, Vincent K. L. Lam⁶⁷, Kyong Soo Park^{68,71}, Danish Saleheen^{72,73}, Wing Yee So⁶⁷, Claudia H. T. Tam⁶⁷, Uzma Afzal⁴⁴, David Aguilar⁷⁴, Rector Arya^{75†}, Tin Aung^{38,39,40}, Edmund Chan⁷⁶, Carmen Navarro^{77,78,79}, Ching-Yu Cheng^{30,38,39,40}, Domenico Palli⁸⁰, Adolfo Correa⁸¹, Joanne E. Curran⁸², Denis Rysin¹⁷, Vidya S. Farook⁸³, Sharon P. Fowler⁴⁹, Barry I. Freedman⁸⁴, Michael Griswold⁸⁵, Daniel Esten Hale⁷⁵, Pamela J. Hicks^{53,54,55}, Chiea-Chuen Khor^{30,38,39,86,87}, Satish Kumar⁸², Benjamin Lehne⁴⁴, Dorothee Thuillier⁸⁸, Wei Yen Lim³⁰, Jianjun Liu^{30,87}, Yvonne T. van der Schouw⁸⁹, Marie Loh^{44,90,91†}, Solomon K. Musani⁹², Sobha Puppala⁸³, William R. Scott⁴⁴, Loic Yengo⁸⁸, Sian-Tsung Tan^{45,61}, Herman A. Taylor Jr⁸¹, Farook Thameem^{49†}, Gregory Wilson Sr⁹³, Tien Yin Wong^{38,39,40}, Pål Rasmus Njølstad^{94,95}, Jonathan C. Levy¹², Massimo Mangino¹¹, Lori L. Bonnycastle¹⁹, Thomas Schwarzmayer⁹⁶, João Fadista⁹⁷, Gabriela L. Surdulescu¹¹, Christian Herder^{98,99}, Christopher J. Groves¹², Thomas Wieland⁹⁶, Jette Bork-Jensen²⁶, Ivan Brandslund^{100,101}, Cramer Christensen¹⁰², Heikki A. Koistinen^{103,104,105,106}, Alex S. F. Doney¹⁰⁷, Leena Kinnunen¹⁰³, Tõnu Esko^{4,108,109,110}, Andrew J. Farmer¹¹¹, Liisa Hakaste^{104,112,113}, Dylan Hodgkiss¹¹, Jasmina Kravic⁹⁷, Valeriya Lysenko⁹⁷, Mette Hollensted²⁶, Marit E. Jørgensen¹¹⁴, Torben Jørgensen^{115,116,117}, Claes Ladvall^{97†}, Johanne Marie Justesen²⁶, Annemari Käräjämäki^{118,119}, Jennifer Kriebel^{99,120,121}, Wolfgang Rathmann¹²², Lars Lannfelt⁵⁸, Torsten Lauritzen¹²³, Narisu Narisu¹⁹, Allan Linneberg^{115,124,125}, Olle Melander¹²⁶, Lili Milani¹⁰⁸, Matt Neville^{12,127}, Marju Orho-Melander¹²⁸, Lu Qi^{129,130†}, Qibin Qi^{129,131}, Michael Roden^{98,99,132}, Olov Rolandsson¹³³, Amy Swift¹⁹, Anders H. Rosengren⁹⁷, Kathleen Styrups¹³⁴, Andrew R. Wood⁹, Evelin Mihailov¹⁰⁸, Christine Blancher¹³⁴, Mauricio O. Carneiro⁴, Jared Maguire⁴, Ryan Poplin⁴, Khalid Shakir⁴, Timothy Fennell⁴, Mark DePristo^{4†}, Martin Hrabé de Angelis^{99,135,136}, Panos Deloukas^{137,138}, Anette P. Gjesing²⁶, Goo Jun¹²⁹, Peter Nilsson¹³⁹, Jacquelyn Murphy⁴, Robert Onofrio⁴, Barbara Thorand^{99,120}, Torben Hansen^{26,140}, Christa Meisinger^{99,120}, Frank B. Hu^{31,129}, Bo Isomaa^{112,141}, Fredrik Karpe^{12,127}, Liming Liang^{20,31}, Annette Peters^{25,99,120}, Cornelia Huth^{99,120}, Stephen P. O'Rahilly¹⁴², Colin N. A. Palmer¹⁴³, Oluf Pedersen²⁶, Rainer Rauramaa¹⁴⁴, Jaakko Tuomilehto^{103,145,146,147,148†}, Veikko Salomaa¹⁴⁸, Richard M. Watanabe^{149,150,151}, Ann-Christine Syvänen¹⁵², Richard N. Bergman¹⁵³, Dwaipayan Bharadwaj^{154†}, Erwin P. Bottinger⁶⁰, Yoon Shin Cho¹⁵⁵, Giriraj R. Chandak^{156†}, Juliana C. N. Chan^{67,157,158}, Kee Seng Chia³⁰, Mark J. Daly⁶³, Shah B. Ebrahim⁵⁹, Claudia Langenberg¹⁰, Paul Elliott^{44,159}, Kathleen A. Jablonski¹⁶⁰, Donna M. Lehman⁴⁹, Weiping Jia³⁷, Ronald C. W. Ma^{67,157,158}, Toni I. Pollán¹⁶¹, Manjinder Sandhu^{13,66}, Nikhil Tandon¹⁶², Philippe Froguel^{88,163}, Inês Barroso^{13,142}, Yik Ying Teo^{30,164,165}, Eleftheria Zeggini¹³, Ruth J. F. Loos⁶⁰, Kerrin S. Small¹¹, Janina S. Ried²², Ralph A. DeFronzo⁴⁹, Harald Grallert^{99,120,121}, Benjamin Glaser¹⁶⁶, Andres Metspalu¹⁰⁸, Nicholas J. Wareham¹⁰, Mark Walker¹⁶⁷, Eric Banks⁴, Christian Gieger^{22,120,121}, Erik Ingelsson^{6,168†}, Hae Kyung Im²⁷, Thomas Illig^{121,169,170}, Paul W. Franks^{35,129,133}, Gemma Buck¹³⁴, Joseph Trakalo¹³⁴, David Buck¹³⁴, Inga Prokopenko^{6,12,163}, Reedik Mägi¹⁰⁸, Lars Lind¹⁷¹, Yossi Farjoun¹⁷², Katharine R. Owen^{12,127}, Anna L. Gloyn^{6,12,127}, Konstantin Strauch^{22,24}, Tiinamaija Tuomi^{104,112,113,173}, Jaspal Singh Kooner^{45,61,174}, Jong-Young Lee^{32†}, Taesung Park^{28,41}, Peter Donnelly^{6,8}, Andrew D. Morris^{175,176}, Andrew T. Hattersley¹⁷⁷, Donald W. Bowden^{53,54,55}, Francis S. Collins¹⁹, Gil Atzmon^{46,178}, John C. Chambers^{44,45,174}, Timothy D. Spector¹¹, Markku Laakso^{51,52}, Tim M. Strom^{96,179}, Graeme I. Bell¹⁸⁰, John Blangero⁸², Ravindranath Duggirala⁸³, E. Shyong Tai^{30,76,181}, Gilean McVean^{6,182}, Craig L. Hanis²⁹, James G. Wilson¹⁸³, Mark Seielstad^{184,185}, Timothy M. Frayling⁹, James B. Meigs¹⁸⁶, Nancy J. Cox^{27†}, Rob Sladek^{15,42,187}, Eric S. Lander¹⁸⁸, Stacey Gabriel⁴, Noël P. Burt⁴, Karen L. Mohlke¹⁸⁹, Thomas Meitinger^{96,179}, Leif Groop^{97,173}, Goncalo Abecasis¹, Jose C. Florez^{4,64,190,191}, Laura J. Scott¹, Andrew P. Morris^{6,108,192}, Hyun Min Kang¹, Michael Boehnke^{1§}, David Altshuler^{4,5,109,190,191,193†§} & Mark I. McCarthy^{6,12,127§}

¹Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. ²Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. ³Center for Biomedicine, European Academy of Bolzano/Bozen (EURAC), affiliated with the University of Lubeck, Bolzano, Italy. ⁴Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. ⁵Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. ⁶Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁷Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ⁸Department of Statistics, University of Oxford, Oxford, UK. ⁹Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. ¹⁰MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. ¹¹Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. ¹²Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ¹³Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK. ¹⁴School of Computer Science, McGill University, Montreal, Quebec, Canada. ¹⁵McGill University and Génomique Québec Innovation Centre, Montreal, Quebec, Canada. ¹⁶Human Genetics Center, The University of Texas Graduate School of Biomedical Sciences at Houston, The University of Texas Health Science Center at Houston, Houston, Texas, USA. ¹⁷Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. ¹⁸National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. ¹⁹Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. ²⁰Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. ²¹Chronic Disease Epidemiology, Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland. ²²Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ²³Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany. ²⁴Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. ²⁵DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. ²⁶The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ²⁷Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, Illinois, USA. ²⁸Department of Statistics, Seoul National University, Seoul, South Korea. ²⁹Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas, USA. ³⁰Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore. ³¹Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. ³²Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do, South Korea. ³³The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA. ³⁴Departments of Computational Medicine & Bioinformatics and Human Genetics, University of Michigan, Ann Arbor, Michigan, USA. ³⁵Department of Clinical Sciences, Lund University Diabetes Centre, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden. ³⁶Department of Epidemiology, Colorado School of Public Health, University of Colorado, Aurora, Colorado, USA. ³⁷Department of Endocrinology and Metabolism, Shanghai Diabetes Institute, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China. ³⁸Singapore Eye Research Institute, Singapore National Eye Centre, Singapore. ³⁹Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore. ⁴⁰The Eye Academic Clinical Programme, Duke-NUS Graduate Medical School, Singapore. ⁴¹Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, South Korea. ⁴²Department of Human Genetics, McGill University, Montreal, Quebec, Canada. ⁴³Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada. ⁴⁴Department of Epidemiology and Biostatistics, Imperial College London, London, UK. ⁴⁵Department of Cardiology, Ealing Hospital NHS Trust, Southall, Middlesex, UK. ⁴⁶Departments of Medicine and Genetics, Albert Einstein College of Medicine, New York, USA. ⁴⁷Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania - Perelman School of Medicine, Philadelphia, Pennsylvania, USA. ⁴⁸Department of Genetics, University of Pennsylvania - Perelman School of Medicine, Philadelphia, Pennsylvania, USA. ⁴⁹Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, USA. ⁵⁰Research, South Texas Veterans Health Care System, San Antonio, Texas, USA. ⁵¹Faculty of Health Sciences, Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. ⁵²Kuopio University Hospital, Kuopio, Finland. ⁵³Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. ⁵⁴Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. ⁵⁵Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. ⁵⁶Centre for Research in Epidemiology and Population Health, Inserm U1018, Villejuif, France. ⁵⁷German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany. ⁵⁸Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden. ⁵⁹Centre for Chronic Disease Control, New Delhi, India. ⁶⁰The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, USA. ⁶¹National Heart and Lung Institute, Cardiovascular Sciences, Hammersmith Campus, Imperial College London, London, UK. ⁶²Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA. ⁶³Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. ⁶⁴Center for Human Genetic

Research, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. ⁶⁵Department of Psychiatry, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA. ⁶⁶Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁶⁷Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China. ⁶⁸Department of Internal Medicine, Seoul National University College of Medicine, Seoul, South Korea. ⁶⁹Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ⁷⁰NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁷¹Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, and College of Medicine, Seoul National University, Seoul, South Korea. ⁷²Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ⁷³Center for Non-Communicable Diseases, Karachi, Pakistan. ⁷⁴Cardiovascular Division, Baylor College of Medicine, Houston, Texas, USA. ⁷⁵Department of Pediatrics, University of Texas Health Science Center, San Antonio, Texas, USA. ⁷⁶Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore. ⁷⁷Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain. ⁷⁸CIBER Epidemiología y Salud Pública (CIBERESP), Universidad de Murcia, Murcia, Spain. ⁷⁹Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain. ⁸⁰Cancer Research and Prevention Institute (ISPO), Florence, Italy. ⁸¹Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, USA. ⁸²South Texas Diabetes and Obesity Institute, Regional Academic Health Center, University of Texas Rio Grande Valley, Brownsville, Texas, USA. ⁸³Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, USA. ⁸⁴Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. ⁸⁵Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, Mississippi, USA. ⁸⁶Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Singapore. ⁸⁷Division of Human Genetics, Genome Institute of Singapore, A*STAR, Singapore. ⁸⁸CNRS-UMR8119, Lille University, Lille Pasteur Institute, Lille, France. ⁸⁹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands. ⁹⁰Institute of Health Sciences, University of Oulu, Oulu, Finland. ⁹¹Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. ⁹²Jackson Heart Study, University of Mississippi Medical Center, Jackson, Mississippi, USA. ⁹³College of Public Services, Jackson State University, Jackson, Mississippi, USA. ⁹⁴KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway. ⁹⁵Department of Pediatrics, Haukeland University Hospital, Bergen, Norway. ⁹⁶Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ⁹⁷Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Malmö, Sweden. ⁹⁸Institute of Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany. ⁹⁹German Center for Diabetes Research (DZD), Neuherberg, Germany. ¹⁰⁰Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. ¹⁰¹Department of Clinical Biochemistry, Vejle Hospital, Vejle, Denmark. ¹⁰²Department of Internal Medicine and Endocrinology, Vejle Hospital, Vejle, Denmark. ¹⁰³Department of Health, National Institute for Health and Welfare, Helsinki, Finland. ¹⁰⁴Abdominal Center: Endocrinology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. ¹⁰⁵Minerva Foundation Institute for Medical Research, Helsinki, Finland. ¹⁰⁶Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. ¹⁰⁷Division of Cardiovascular and Diabetes Medicine, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, UK. ¹⁰⁸Estonian Genome Center, University of Tartu, Tartu, Estonia. ¹⁰⁹Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. ¹¹⁰Division of Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA. ¹¹¹Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. ¹¹²Folkhälsan Research Centre, Helsinki, Finland. ¹¹³Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland. ¹¹⁴Steno Diabetes Center, Gentofte, Denmark. ¹¹⁵Research Centre for Prevention and Health, Capital Region of Denmark, Glostrup, Denmark. ¹¹⁶Department of Public Health, Institute of Health Sciences, University of Copenhagen, Copenhagen, Denmark. ¹¹⁷Faculty of Medicine, Aalborg University, Aalborg, Denmark. ¹¹⁸Department of Primary Health Care, Vaasa Central Hospital, Vaasa, Finland. ¹¹⁹Diabetes Center, Vaasa Health Care Center, Vaasa, Finland. ¹²⁰Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ¹²¹Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ¹²²Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany. ¹²³Department of Public Health, Section of General Practice, Aarhus University, Aarhus, Denmark. ¹²⁴Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark. ¹²⁵Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹²⁶Department of Clinical Sciences, Hypertension and Cardiovascular Disease, Lund University, Malmö, Sweden. ¹²⁷Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. ¹²⁸Department of Clinical Sciences, Diabetes and Cardiovascular Disease, Genetic Epidemiology, Lund University, Malmö, Sweden. ¹²⁹Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. ¹³⁰Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. ¹³¹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, USA. ¹³²Department of Endocrinology and Diabetology, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany. ¹³³Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. ¹³⁴High Throughput Genomics, Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹³⁵Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ¹³⁶Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising-Weihenstephan, Germany. ¹³⁷William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ¹³⁸Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. ¹³⁹Department of Clinical Sciences, Medicine, Lund University, Malmö, Sweden. ¹⁴⁰Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. ¹⁴¹Department of Social Services and Health Care, Jakobstad, Finland. ¹⁴²Metabolic Research Laboratories, Institute of Metabolic Science, University of Cambridge, Cambridge, UK. ¹⁴³Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK. ¹⁴⁴Foundation for Research in Health, Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. ¹⁴⁵Center for Vascular Prevention, Danube University Krems, Krems, Austria. ¹⁴⁶Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. ¹⁴⁷Instituto de Investigación Sanitaria del Hospital Universitario La Paz (IdiPAZ), University Hospital La Paz, Autonomous University of Madrid, Madrid, Spain. ¹⁴⁸National Institute for Health and Welfare, Helsinki, Finland. ¹⁴⁹Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ¹⁵⁰Department of Physiology & Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ¹⁵¹Diabetes and Obesity Research Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ¹⁵²Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ¹⁵³Cedars-Sinai Diabetes and Obesity Research Institute, Los Angeles, California, USA. ¹⁵⁴Functional Genomics Unit, CSIR-Institute of Genomics & Integrative Biology (CSIR-IGIB), New Delhi, India. ¹⁵⁵Department of Biomedical Science, Hallym University, Chuncheon, South Korea. ¹⁵⁶CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India. ¹⁵⁷Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, China. ¹⁵⁸Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, China. ¹⁵⁹MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. ¹⁶⁰The Biostatistics Center, The George Washington University, Rockville, Maryland, USA. ¹⁶¹Department of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. ¹⁶²Department of Endocrinology and Metabolism, All India Institute of Medical Sciences, New Delhi, India. ¹⁶³Department of Genomics of Common Disease, School of Public Health, Imperial College London, London, UK. ¹⁶⁴Life Sciences Institute, National University of Singapore, Singapore. ¹⁶⁵Department of Statistics and Applied Probability, National University of Singapore, Singapore. ¹⁶⁶Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel. ¹⁶⁷The Medical School, Institute of Cellular Medicine, Newcastle University, Newcastle, UK. ¹⁶⁸Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. ¹⁶⁹Hannover Unified Biobank, Hannover Medical School, Hanover, Germany. ¹⁷⁰Institute for Human Genetics, Hannover Medical School, Hanover, Germany. ¹⁷¹Department of Medical Sciences, Uppsala University, Uppsala, Sweden. ¹⁷²Data Sciences and Data Engineering, Broad Institute, Cambridge, Massachusetts, USA. ¹⁷³Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. ¹⁷⁴Imperial College Healthcare NHS Trust, Imperial College London, London, UK. ¹⁷⁵Clinical Research Centre, Centre for Molecular Medicine, Ninewells Hospital and Medical School, Dundee, UK. ¹⁷⁶The Usher Institute to the Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ¹⁷⁷University of Exeter Medical School, University of Exeter, Exeter, UK. ¹⁷⁸Department of Natural Science, University of Haifa, Haifa, Israel. ¹⁷⁹Institute of Human Genetics, Technische Universität München, Munich, Germany. ¹⁸⁰Departments of Medicine and Human Genetics, The University of Chicago, Chicago, Illinois, USA. ¹⁸¹Cardiovascular & Metabolic Disorders Program, Duke-NUS Medical School Singapore, Singapore. ¹⁸²Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. ¹⁸³Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, USA. ¹⁸⁴Department of Laboratory Medicine & Institute for Human Genetics, University of California, San Francisco, San Francisco, California, USA. ¹⁸⁵Blood Systems Research Institute, San Francisco, California, USA. ¹⁸⁶General Medicine Division, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. ¹⁸⁷Division of Endocrinology and Metabolism, Department of Medicine, McGill University, Montreal, Quebec, Canada. ¹⁸⁸Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ¹⁸⁹Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. ¹⁹⁰Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. ¹⁹¹Diabetes Research Center (Diabetes Unit), Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. ¹⁹²Department of Biostatistics, University of Liverpool, Liverpool, UK. ¹⁹³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

*These authors contributed equally to this work.

‡These authors jointly supervised this work.

§Present addresses are provided in the Supplementary Information.

‡Deceased.

METHODS

Ethics statement. All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Data generation

GoT2D integrated panel generation

GoT2D sequenced samples. Here we describe how we generated, processed, and carried out quality control (QC) on sequence and genotype data for the 2,891 individuals initially chosen for GoT2D from four studies, and how this resulted in 2,657 individuals (1,326 T2D cases and 1,331 non-diabetic controls) for analysis (Extended Data Fig. 1). We preferentially sampled early-onset, lean, and/or familial T2D cases and overweight controls with low fasting glucose levels⁵⁰. Specific details of selected samples are provided in Extended Data Table 1 and Supplementary Table 1.

DNA sample preparation. De-identified DNA samples were sent to the Broad Institute (DGI, FUSION), Wellcome Trust Centre for Human Genetics in Oxford (UKT2D), and Helmholtz Zentrum München (KORA) and prepared for genetic analysis. DNA quantity was measured by Picogreen (all), and samples with sufficient total DNA and minimum concentrations for downstream experiments were genotyped for a set of 24 SNVs using the Sequenom iPLEX assay (DGI, FUSION, UKT2D); one gender assay and 23 SNVs located across the autosomes. The genotypes for these SNVs were used as a quality filter to advance samples and a technical fingerprint for subsequent sequencing and genome-wide array genotypes.

Exome sequencing. Genomic DNA was sheared, end repaired, ligated with barcoded Illumina sequencing adapters, amplified, size selected, and subjected to in-solution hybrid capture using the Agilent SureSelect Human All Exon 44Mb v2.0 (DGI, FUSION, UKT2D) and v3.0 (KORA) bait set (Agilent Technologies, USA). Resulting Illumina exome sequencing libraries were qPCR quantified, pooled, and sequenced with 76-bp paired-end reads using Illumina GAI or HiSeq 2000 sequencers to ~82-fold mean coverage.

Genome sequencing. Whole-genome Illumina sequencing library construction was performed as described for exome capture above, except that genomic DNA was sheared to a larger target size and hybrid capture was not performed. The resulting libraries were size selected to contain fragment insert size of 380 bp \pm 20% (DGI, FUSION, KORA) and 420 bp \pm 25% (UKT2D) using gel electrophoresis or the SAGE Pippin Prep (Sage Science, USA). Libraries were qPCR quantified, pooled, and sequenced with 101-bp paired-end reads using Illumina GAI or HiSeq 2000 sequencers to ~5-fold mean coverage.

HumanOmni2.5 array genotyping. Genotyping was performed by the Broad Genetic Analysis Platform. DNA samples were placed on 96-well plates and genotyped using the Illumina HumanOmni2.5-4v1_B SNV array.

Alignment and processing of exome and genome sequence data

Alignment of sequence reads to reference genome. Sequence data were processed and aligned to hg19 using the Picard (<http://broadinstitute.github.io/picard/>), BWA⁵¹, and GATK^{52,53} pipelines. The resulting BAM and VCF files were submitted to NCBI and are available in dbGaP (accession number phs000840.v1.p1, study name NIDDK_GoT2D).

Coverage and QC of aligned sequence reads. We excluded 151 exome samples with average coverage $\leq 20\times$ in $>20\%$ of the target bases and 68 genome samples with average coverage $\leq 5\times$. After sequence alignment and post-processing, aligned sequence reads were screened based on multiple QC criteria, including number of mapped reads, number of mapped bases with $<1\%$ estimated base call error rate ($>Q20$), fraction of duplicate reads, fraction of properly paired reads, distribution of insert sizes, distribution of mean base quality with respect to sequencing cycles, and GC bias (Extended Data Fig. 1).

Detecting and handling contamination of sequence reads. We assessed possible DNA contamination in the genome and exome sequence data using verifyBamID⁵⁴ using two methods. First, we estimated the contamination level of sequenced samples using allele frequencies estimated from the HumanOmni2.5 array on a thinned set of 100,000 markers with minor allele frequency (MAF) $> 5\%$. Second, for samples with HumanOmni2.5 genotypes, we used these genotypes together with sequence data to estimate contamination and identify possible sample swaps. We excluded exome sequence data for 7 individuals and genome sequence data for 59 individuals with estimated contamination $\geq 2\%$ using either method. Prior to variant calling, uncontaminated sample swaps were assigned to the correct sample label after searching for the matching pairs using the same method.

GoT2D integrated panel genotype calling

SNV identification. We processed whole-genome sequence reads across the remaining 2,764 QC-passed individuals by two SNV calling pipelines: GotCloud (www.gotcloud.org) and GATK UnifiedGenotyper⁵⁵. We merged unfiltered SNV calls across the two call sets and then processed the merged site list through the SVM and VQSR filtering algorithms implemented by those pipelines. SNVs that failed both filtering algorithms were removed before genotyping and haplotype integration. For the 2,733 QC-passed exome-sequenced individuals, we used GATK UnifiedGenotyper to call SNVs.

Illumina HumanOmni2.5 array genotyping. We used Illumina GenomeStudio v2010.3 with default clusters to call HumanOmni2.5 genotypes after comparing different clustering algorithms and observing that the default cluster resulted in highest concordance with sequence-based genotypes. Called genotypes were run through a standard QC pipeline; samples passing a call rate threshold of 95%, and genetic fingerprint (24-marker panel) and gender concordance were passed on to downstream GWAS QC. SNVs with GenTrain score < 0.6 , cluster separation score < 0.4 , or call rate $< 97\%$ were considered technical failures at the genotyping laboratory and deleted before data release. We removed samples with call rate $< 98\%$, and SNVs monomorphic across all samples, failed by 1000G Omni 2.5 QC filter, or with Hardy–Weinberg equilibrium $P < 10^{-6}$ (Extended Data Fig. 1). Eight-five samples were removed in this process.

Short insertion and deletion (indel) identification. For the whole-genome sequence data, we used the GATK UnifiedGenotyper to call short indels (< 50 bp). Because short indels are known to have high false positive rates due to systematic sequencing and alignment errors⁵⁵, we used stringent filtering criteria in SVM and VQSR and excluded indels that failed either algorithm. For exome sequencing, we used GATK UnifiedGenotyper to call short indels, following best practices described elsewhere⁵².

Large deletion identification. We used GenomeSTRIP⁵⁶ to call large (> 100 -bp) deletions in the whole-genome sequence data. After initial discovery of large deletions in 2,764 QC-passed individuals, we merged the discovered sites with deletions identified in 1,092 sequenced individuals from the 1000G Project to increase sensitivity and then genotyped the merged site lists across the 2,764 individuals. After applying the default filtering implemented in GenomeSTRIP, pass-filtered sites variable in any of the samples were identified as candidate variant sites. Among these candidate sites, we excluded variants in known immunoglobulin loci to reduce the impact of possible cell-line artefacts. We then excluded 136 more individuals owing to an unusually large number of variants per sample ($> \text{median} + (3 \times \text{mean absolute deviation})$). Variants present only in these excluded individuals were removed from further analysis.

GoT2D integrated panel haplotype integration

Genotype likelihood calculation. We merged SNVs discovered from the three experimental platforms into one site list and calculated genotype likelihoods across all sites separately by platform. Because exome sequence data have substantial off-target coverage, we calculated likelihoods across the genome combining data from the genome and exome sequence experiments. For genome sequence, we calculated likelihoods using GotCloud; for exomes, we used GATK UnifiedGenotyper; for HumanOmni2.5 genotypes, we converted hard genotype calls into genotype likelihoods assuming a genotype error rate of 10^{-6} . For indels, we calculated likelihoods in a similar way except that the HumanOmni2.5 data could not be used. For structural variants (SVs), genotype likelihoods were calculated from GenomeSTRIP using the whole-genome sequence data.

Integration of genotype and sequence data. We calculated combined genotype likelihoods across each of the 2,874 individuals as the product of the corresponding genome, exome, and HumanOmni2.5 likelihoods assuming independent data across platforms (Extended Data Fig. 1). We then phased the genotype data using the strategy developed for 1000G Phase 1 (ref. 55). Specifically, we phased the integrated likelihoods using Beagle⁵⁷ with 10,000 SNVs per chunk and 1,000 overlapping SNVs between consecutive chunks. We refined phased sequences using Thunder⁵⁸ as implemented in GotCloud (<http://genome.sph.umich.edu/wiki/GotCloud>) with 400 states to improve genotype and haplotype quality.

GoT2D integrated panel QC. 2,874 individuals were available in the integrated haplotype panel. To identify population outliers, we carried out principal components analysis (PCA). We computed PCs for each of the three variant types (SNVs, short indels, large deletions) using EPACTS on an LD-pruned ($r^2 < 0.20$) set of autosomal variants obtained by removing large high-LD regions^{59,60}, variants with MAF < 0.01 , and variants with Hardy–Weinberg equilibrium $P < 10^{-6}$. Inspecting the first ten PCs for each variant type, we identified 43 population outliers and 136 additional outliers for large deletions only; we excluded these 179 individuals. We excluded an additional 38 individuals based on close relationships (estimated genome-wide identity-by-descent proportion of alleles shared > 0.20) with other study members. 2,657 individuals remained available for downstream analyses (Extended Data Fig. 1).

GoT2D integrated panel evaluation of variant detection sensitivity. As we had no external data to evaluate SNV and indel variant detection sensitivity and genotype accuracy for our integrated haplotype panel, we evaluated accuracy for the low-pass whole-genome sequence data using the exome sequence data as a gold standard for variants at which exome sequence depth was ≥ 10 . We consider the resulting sensitivity and accuracy estimates as lower bounds for the integrated panel, which combined information from the genome, exome, and HumanOmni2.5 data.

We estimated the sensitivity of low-pass genome sequence data to detect true SNVs by calculating the proportion of exome-sequencing-detected SNVs detected by low-pass genome sequencing in the 2,538 individuals with data for all three experimental platforms. For exome sequence allele counts $<1,000$, we merged adjacent allele count bins until the number of alleles was $>1,000$. We estimated the sensitivity of low-pass genome sequencing to detect common, low-frequency, and rare SNVs as 99.8%, 99.0%, and 48.2%, respectively. Similarly, we estimated the sensitivity of low-pass genome sequence to detect true short indels by calculating the proportion of exome sequencing-detected short indels detected by low-pass genome sequencing. Sensitivity estimates were $>99.9\%$, 93.8%, and 17.9% for common, low-frequency, and rare short indels, respectively.

To estimate the sensitivity of the combined low-pass genome and exome sequence data, we focused on coding SNVs and calculated the proportion of HumanOmni2.5 SNVs detected by either sequencing platform. Because HumanOmni2.5 SNVs are enriched for common variants, we calculated a weighted averaged sensitivity at each allele count, weighted by the number of exome-detected variants given the allele count. Sensitivity estimates were 99.9%, 99.7%, and 83.9% for common, low-frequency, and rare variants.

GoT2D integrated panel evaluation of genotype accuracy. To evaluate genotype accuracy for SNVs, we focused on chromosome 20, and compared the concordance of low-pass whole-genome-sequence-based genotypes with those based on exome sequence. Overall genotype concordance was 99.86%. Homozygous reference, heterozygous, and homozygous non-reference concordances were 99.97%, 98.34%, and 99.72%, respectively. We also compared genotype concordance between exome sequence and HumanOmni2.5 genotypes. Overall concordance was 99.4%. When the HumanOmni2.5 genotypes were homozygous reference, heterozygous, and homozygous non-reference, concordances were 99.97%, 99.69%, and 99.88%, respectively.

We evaluated the genotype accuracy of indels for the 210 chromosome-20 indels that overlapped between those discovered by exome and genome sequencing. Overall genotype concordance was 99.4%. When the exome genotypes were homozygous reference, heterozygous, and homozygous non-reference, concordances were 99.8%, 95.8%, and 98.6%, respectively.

To evaluate the genotype accuracy of our low-pass genome sequence data to detect true structural variants, we took advantage of the 181 individuals in our study who were previously included in the WTCCC array-CGH based structural variant detection experiment⁶¹. Taking the WTCCC data as a gold standard, we estimated genotype accuracy across 1,047 overlapping structural variants (with reciprocal overlap >0.8) genome-wide. The overall genotype concordance was 99.8%. When the WTCCC genotypes were homozygous reference, heterozygous, and homozygous non-reference, concordances were 99.9%, 99.6%, and 99.7%, respectively.

GoT2D+T2D-GENES multiethnic exome panel generation and QC

Samples. We considered 6,504 T2D cases and 6,436 controls from 14 studies of African American, East Asian, South Asian, Hispanic, and European ancestry. In contrast to the GoT2D whole-genome integrated panel, this data set also includes GoT2D individuals for whom whole-genome data were not available. Sample characteristics are provided in Extended Data Table 1 and Supplementary Table 4.

Sequence reads were processed and aligned to the reference genome (hg19) with Picard (<http://broadinstitute.github.io/picard/>). Polymorphic sites and genotypes were called with GATK, with filtering of sites performed using Variant Quality Score Recalibration (VSR) for SNVs, and hard filters for indels. Genotype likelihoods were computed controlling for contamination.

Hard calls (the GATK-called genotypes but set as missing at a genotype quality (GQ) <20 threshold⁵²) and dosages (the expected value of the genotype, defined as $Pr(RX|data) + 2Pr(XX|data)$, where X is the alternative allele) were computed for each sample at each variant site. Hard calls were used only for quality control, while dosages were used in all downstream association analyses. Multi-allelic SNVs and indels were dichotomized by collapsing alternate alleles into one category because downstream association analyses required bi-allelic variants.

Individuals were excluded from analysis if they were outliers on one of multiple metrics: poor array genotype concordance (where available), high number of variant alleles or singletons, high or low allele balance (average proportion of non-reference alleles at heterozygous sites), or excess mean heterozygosity or ratio of heterozygous to homozygous genotypes.

Within this reduced set of individuals, we then performed extended QC using ethnicity and T2D status to provide high-quality genotype data for downstream association analyses. Within each ethnicity, we excluded variants based on hard call rate ($<90\%$ in any cohort), deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$ in any ancestry group), or differential call rate between T2D cases and controls ($P < 10^{-4}$ in any ancestry group). We then considered autosomal variants that passed extended QC and with MAF $>1\%$ in all ancestry groups for trans-ethnic kinship analyses. We calculated identity-by-state (IBS) between each pair of

samples based on independent variants (trans-ethnic $r^2 < 0.05$) and constructed axes of genetic variation through PCA implemented in EIGENSTRAT⁶² to identify ethnic outliers (Supplementary Fig. 29). We also identified duplicates based on IBS, and excluded the sample from each pair with lowest call rate and/or mismatch with external information. The extended QC excluded 68 individuals, 9.9% of SNVs and 90.8% of indels from the clean data set.

Association analysis

Power calculation. We used the genetic power calculator (<http://pngu.mgh.harvard.edu/~purcell/gpc/>) to estimate power to detect T2D association assuming 8% prevalence. For the T2D-GENES+GoT2D exome sequence data set we assumed: (i) a fixed-effect across all five ancestry groups (12,940 individuals); and (ii) an effect specific to one group (2,000 individuals) (Extended Data Fig. 2). We repeated our calculations for combined exome sequence and exome array data, assuming a fixed effect across all ethnicities, for an effective total sample size of 82,758 individuals (Extended Data Fig. 2).

For the GoT2D integrated panel we allowed for incomplete variant detection by multiplying power by the estimated sensitivity to detect the variant as a function of MAF. For imputed variants, we first multiplied the sample size by the median imputation quality (rsq_hat) obtained from MaCH/Thunder or minimac⁶³ for the corresponding MAF bin across the analysed cohorts, and then multiplied the estimated power by the fraction of variants that passed the imputation quality cutoff for that MAF bin.

For gene-based tests in the T2D-GENES+GoT2D data, we made use of a Bonferroni correction for 20,000 genes, corresponding to $P < 2.5 \times 10^{-6}$. We used a simulated haplotype data set from the SKAT package (<http://cran.r-project.org/web/packages/SKAT/vignettes/SKAT.pdf>) and estimated the power of SKAT-O to detect association of variants within a gene at this threshold as a function of the phenotypic variance (1%) in a liability scale explained by additive genetic effects and the percentage of variants that were causal (50% and 100%). As for single-variant power calculations, we considered: (i) a fixed-effect across all ethnicities (12,940 individuals); and (ii) an effect specific to one ancestry group (2,000 individuals) (Extended Data Fig. 2).

GoT2D integrated panel association analysis

Single-variant association analysis. We tested for T2D association in a logistic regression framework assuming an additive genetic model. We used the Firth bias-corrected likelihood ratio test^{64,65} as our primary analysis strategy; we repeated association analysis using the score test for inclusion in sample-size-weighted meta-analysis (Supplementary Table 2). Tests were adjusted for sex, the first two genotype-based PCs to account for population stratification, and an indicator function for observed temporal stratification based on sequencing date and centre. PCs were calculated using linkage-disequilibrium (LD) pruned ($r^2 < 0.20$) HumanOmni2.5M array variants with MAF $>1\%$ after removing large high-LD regions^{59,60}.

Aggregate association analysis. To test for aggregate association within coding regions of the genome, we used the approach described in *Gene-based analysis* below. For every gene and mask tested, P values were greater than 2.5×10^{-4} .

We also tested for aggregate association among variants in non-coding regions of the genome. We aggregated variants in individual pancreatic islet enhancer elements (see *Genomic annotation* below), as these elements collectively demonstrated strongest genome-wide enrichment of T2D association. We performed both the burden and SKAT tests using genotypes from the integrated panel on variants with MAF $<5\%$ in each islet enhancer element. We used a Bonferroni threshold $P < 1.68 \times 10^{-7}$ based on a nominal significance level of $\alpha = 0.05$ corrected for 298,240 elements with at least one variant. All elements tested in this manner had P values greater than 2.5×10^{-6} .

GoT2D+T2D-GENES multiethnic association analysis

Kinship analysis. Within each ancestry group, we considered autosomal variants that passed QC with MAF $>1\%$ for ethnic-specific kinship analyses. We calculated IBS between each pair of samples in the ancestry group based on independent variants (ethnic-specific $r^2 < 0.05$) and constructed a kinship matrix to account for intra-ethnic population structure and relatedness in downstream mixed-model (EMMAX)-based association analyses¹⁶. We also used IBS to identify pairs of related individuals within each ancestry group (defined by $\pi_i\text{-hat} > 0.3$). We then defined intra-ethnic related exclusion lists for downstream non-EMMAX association analyses using the following steps: (i) remove the control from each T2D-status discordant pair; and (ii) remove the sample with lowest call rate from each T2D-status concordant pair. We also constructed intra-ethnic axes of genetic variation through PCA implemented in EIGENSTRAT⁶². We identified axes of genetic variation in each ancestry group for inclusion as covariates in downstream non-EMMAX association analyses to account for intra-ethnic population structure that: (i) explain at least 0.5% genotypic variation; and/or (ii) demonstrate nominal association ($P < 0.05$) with T2D in logistic regression analysis.

Single-variant association analysis. Within each ancestry group, we performed a score test of T2D association with each variant passing ethnic-specific QC in a linear regression framework under an additive model in EMMAX¹⁶. We also performed a Wald test of T2D association with each variant passing ethnic-specific QC in a logistic regression framework under an additive model with adjustment for ethnic-specific axes of genetic variation after exclusion of related samples (Supplementary Table 30). Within each ancestry group, we calculated genomic control inflation factors (score EMMAX and Wald) based on independent variants used for the ethnic-specific kinship analyses and corrected association summary statistics (*P* value and s.e.) to account for residual population structure.

Subsequently, we performed trans-ethnic fixed-effects meta-analysis of ancestry-specific association summary statistics at each variant based on: (i) sample size weighting of score EMMAX directed *P* values; and (ii) inverse-variance weighting of Wald beta/s.e. (to obtain unbiased estimates of allelic odds ratios and confidence intervals that cannot be constructed from EMMAX effect estimates). We also performed trans-ethnic meta-analysis of ancestry-specific association summary statistics (score EMMAX beta/s.e.) at each variant using MANTRA⁶⁶, using pair-wise mean allele frequency differences at the subset of independent variants used for trans-ethnic kinship analyses as a prior for relatedness between ancestry groups.

Validation of PAX4 association signal in additional East Asian studies. We validated the PAX4 Arg192His (rs2233580) association signal in an additional 1,789 T2D cases and 1,509 controls of East Asian ancestry from Hong Kong, Korea, and Singapore (Supplementary Table 9). Within each study, we tested for association with T2D in a logistic regression model, and combined association summary statistics across studies through fixed-effects meta-analysis (Supplementary Table 9). Among T2D cases, we also tested for association with age of diagnosis in a linear regression model, and combined association summary statistics across studies through fixed-effects meta-analysis (Supplementary Table 9).

Admixture analysis. Admixed populations can offer greater statistical power to detect association because diverse ancestry increases genetic variation. However, admixture can also introduce false-positive signals due to population stratification and heterogeneity of effects because of differential LD⁶⁷. To assess the contribution of ancestral background in the two admixed groups (African American and Hispanic), we inferred local ancestry based on SNVs in available GWAS data using two approaches. For African Americans, we ran HAPMIX⁶⁸ using CEU and YRI haplotypes from HapMap as reference, and estimated the proportion of European ancestry at each genomic position. For Hispanics, we ran Multimix⁶⁹ using European, West African, and Native American haplotypes from HapMap as reference, and estimated the proportion of European ancestry at each genomic position, since we observe only a very low West African contribution (1.1–3.2%, Supplementary Fig. 31). We then repeated our intra-ethnic EMMAX-based analyses within African American and Hispanic ancestry groups, this time adjusting for local ancestry by including the estimated proportion of European ancestry at each variant as a covariate. Adjustment for local ancestry resulted in numerically similar association statistics as those from unadjusted analyses in the African American and Hispanic samples.

Gene-based analysis. We generated four variant lists ('masks') based on MAF and functional annotation. We mapped variants to transcripts in Ensembl 66 (GRCh37.66). Using annotations from CHAoS v0.6.3, SnpEFF v3.1, and VEP v2.7, we identified variants predicted to be protein-truncating (for example, nonsense, frameshift, essential splice site) denoted PTV-only or 'Mask 1'; or protein-altering (for example, missense, in-frame indel, non-essential splice site) in at least one mapped transcript (by at least one of the three algorithms) with MAF < 1%, denoted PTV+missense or 'Mask 2'. We additionally used the procedure described by Purcell *et al.*⁷⁰ to identify subsets of missense variants with MAF < 1% meeting 'strict' or 'broad' criteria for being deleterious, using annotation predictions from Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, Mutation Taster, and SIFT; variants predicted deleterious by all five algorithms or by at least one algorithm were denoted PTV+NS_{strict} or 'Mask 3' and PTV+NS_{broad} or 'Mask 4', respectively. Indels predicted by CHAoS, SnpEFF, or VEP to introduce frameshifts were included in the 'strict' category. We calculated MAFs for each ancestry using high-quality genotype calls (GQ > 20) for all samples passing extended QC. We considered a variant to have MAF < 1% if MAF estimates for every ancestry group were < 1%.

We used the MetaSKAT R package (v0.32)¹⁵ with the SKAT v0.93 library to perform SKAT-O⁷¹ analysis within each ancestry, and in meta-analysis. Within each ancestry group, we analysed genotype dosages with adjustment for ethnic-specific axes of genetic variation after exclusion of 96 related individuals. We assumed homogenous allele frequencies and genetic effects for all studies within an ancestry group. We performed meta-analysis using genotype-level data, allowing for heterogeneity of allele frequencies and genetic effects between (but homogeneity

within) ancestry groups. All analyses were completed using the recommended rho vector for SKAT-O: (0, 0.12, 0.22, 0.32, 0.52, 0.5, 1).

Imputed data

Samples. We carried out genotype imputation into 44,414 individuals (11,645 T2D cases and 32,769 controls) from 13 studies using the GoT2D integrated haplotypes as reference panel. Characteristics of the imputed studies are provided in Extended Data Table 1 and Supplementary Table 3.

Single-variant association meta-analysis. The one sequenced and thirteen imputed studies totalled 12,971 T2D cases and 34,100 controls. Each study performed its own sample- and variant-based QC. In each study, SNVs with minor allele count (MAC) ≥ 1 passing QC were tested for T2D association assuming an additive genetic model adjusting for study-specific covariates. Association testing was performed using logistic regression Firth bias-corrected, likelihood ratio, or score tests as implemented in EPACTS (genome.sph.umich.edu/wiki/EPACTS) or SNPTEST⁷². To account for related samples in the Framingham Heart Study, generalized estimating equations (GEE) were used, as implemented in R. Residual population stratification for each study was accounted for using genomic control⁷³. We then carried out fixed-effects sample-size weighted meta-analysis as implemented in METAL⁷⁴.

Conditional analyses in established GWAS loci. We compiled a list of 143 previously reported genome-wide significant SNVs in 81 T2D autosomal loci (i) from Morris *et al.*² and Voight *et al.*⁴; (ii) from papers they referenced; and (iii) from references in the NHGRI GWAS catalogue⁷⁵. We LD pruned these SNVs (*r*² < 0.95), yielding a list of 129 SNVs. We deleted the *CILP2* locus (and two SNVs) from subsequent whole-genome analyses owing to large regions in which no variants passed QC, resulting in a list of 127 index SNVs at 80 autosomal loci. To identify additional T2D-associated variants within these 80 T2D autosomal loci in the genome-wide data, we repeated GWA analysis for 12 of the 13 studies (conditional analysis results for FHS were unavailable), conditioning on the 127 index SNVs. We performed fixed-effects inverse-variance meta-analysis to combine conditional analysis results from the studies totalling 12,298 cases and 26,440 controls. For each known locus, we analysed all SNVs within 500 kb of the known index SNVs; if there were multiple known index SNVs, we analysed all SNVs within 500 kb of the most proximal and distal index SNVs. We imposed a conditional-analysis significance threshold of $\alpha = 1.8 \times 10^{-6}$ based on a proportional number of multiple tests for ~83 Mb of the ~3,000-Mb genome.

Exome array data

Samples. We considered 28,305 T2D cases and 51,549 controls from 13 studies of European ancestry, genotyped with the Illumina exome array. Characteristics of the studies are provided in Extended Data Table 1 and Supplementary Table 15.

Overlap of exome sequence variation with exome array. We assessed overlap of variants present on the exome array with those observed in our trans-ethnic exome-sequence data. As exome array primarily contains SNVs that are predicted to be protein altering, we focused on nonsense, essential splice site, and missense variants. Only variants passing QC in both sequence and array data were included in our overlap assessment.

Data processing, QC, and kinship analysis. Within each study, exome array genotypes were initially called using Illumina GenCall (http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf) and Birdseed⁷⁶. Sample and variant QC was then undertaken within each study based on several quality control filters. Criteria for sample exclusion included low call rate (<99%), mean heterozygosity, high singleton counts, non-European ancestry, sex discrepancy, GWAS discordance (where data were available), genotyping platform fingerprint discordance, and duplicate discordance. Variants were excluded based on call rate (<99%), deviation from Hardy–Weinberg equilibrium ($P < 10^{-6}$), duplicate, chromosome or allele mismatch, GenTrain score < 0.6, cluster separation score < 0.4, and manual cluster checks. Missing genotypes were subsequently re-called using zCall, with a second round of QC to exclude poor quality samples (call rate < 99% and mean heterozygosity) and variants (call rate < 99%). Within each study, we considered independent autosomal variants that passed QC with MAF > 1% for kinship analyses, and calculated IBS between each pair of samples. We used these statistics to: (i) identify non-European ancestry samples to be excluded from all downstream analyses; (ii) construct a kinship matrix to account for fine-scale population structure and relatedness in downstream EMMAX-based association analyses; (iii) identify related samples to be excluded from downstream non-EMMAX association analyses; and (iv) calculate axes of genetic variation for inclusion as covariates in downstream non-EMMAX association analyses to account for fine-scale population structure (if required).

Single-variant association analysis. Within each study, we performed a score test of T2D association with each variant passing QC in a mixed-model regression framework under an additive model in EMMAX¹⁶. We also performed a Wald test of T2D association with each variant in a logistic regression framework under an additive model with adjustment for axes of genetic variation after exclusion of

related samples. For each test, we corrected s.e. and P value for the genomic control inflation factor (if > 1) calculated on the basis of the independent autosomal variants used for kinship analysis.

Across studies, we performed fixed-effects meta-analysis of association summary statistics at each variant based on: (i) inverse-variance weighting of score EMMAX beta/s.e.; (ii) sample size weighting of score EMMAX directed P values; and (iii) inverse-variance weighting of Wald beta/s.e. For each of these meta-analyses, we applied a second round of correction of s.e. and P value by genomic control, again calculated based on the independent autosomal SNVs used for kinship analyses.

Combined exome sequence and exome array single-variant analysis. We considered variants that were represented both in the exome sequence and on the exome chip. We began by performing fixed-effects meta-analysis of association summary statistics (after correction for genomic control, as described above) from the exome-chip meta-analysis and the European ancestry sequenced samples using: (i) inverse-variance weighting of score EMMAX beta/s.e.; (ii) sample size weighting of score EMMAX directed P values; and (iii) inverse-variance weighting of Wald beta/s.e. Subsequently, we performed trans-ethnic fixed-effects meta-analysis of ancestry-specific association summary statistics (after correction for genomic control, as described above) at each variant based on: (i) sample size weighting of score EMMAX directed P values; and (ii) inverse-variance weighting of Wald beta/s.e.

Gene-based analyses. We made use of the four variant masks defined for exome sequence gene-based analyses, but with MAF calculated across all exome array studies. Within each study, we performed SKAT-O analyses⁷¹, with adjustment for axes of genetic variation after exclusion of related samples. We combined P values for association across studies via meta-analysis with Stouffer's method⁷⁷.

Evaluating relationships between association signals for coding variants and previously reported lead SNVs at established GWAS loci. For coding variants that mapped to established T2D susceptibility loci and achieved genome-wide significance in combined exome sequence and/or exome array analysis, we used complementary approaches with a range of available genetic data resources to evaluate their contribution to the association signals of previously reported lead SNVs. If the previously reported lead SNV (or a good proxy, $r^2 \geq 0.8$) was genotyped on the exome array, we performed reciprocal conditional analyses with the available exome array data. Within each study, we repeated EMMAX analyses in GWAS loci, including additively coded genotypes at the previously reported² lead SNV or genome-wide significant coding variant as an additional covariate in the regression model. Across studies, we performed fixed-effects meta-analysis of association summary statistics at each variant based on: (i) inverse-variance weighting of score EMMAX beta/s.e.; (ii) sample size weighting of score EMMAX-directed P values. If the previously reported lead SNV (or a good proxy) was not genotyped on the exome array, we performed approximate reciprocal conditional analysis, implemented in GCTA⁷⁸, using genome-wide meta-analysis association summary statistics from 12,971 T2D cases and 34,100 controls from the combined GoT2D integrated panel and imputed data. Patterns of LD between variants were estimated using a subset of the GoT2D integrated panel, restricted to 2,389 individuals with pairwise genetic relationship < 0.025 , as defined by the GCTA A statistic⁷⁹. Finally, we interrogated 99% credible sets of variants at each GWAS locus, which together represent $\geq 99\%$ of the probability of driving each association signal. We determined whether the coding variant at each locus was included in the credible set for the association signal for the previously reported lead SNV, and recorded its rank.

Enrichment of exome association signals in GWAS. To define T2D-associated intervals, we first identified all SNVs associated with T2D in published genome-wide association studies (GWAS) by searching literature and the NHGRI GWAS catalogue (see also *Conditional analyses in established GWAS loci* above). We identified 143 autosomal SNVs, with some associated in more than one ancestry (167 SNV-ancestry pairs). For each SNV-ancestry pair, we identified the most distant pair of SNVs with $r^2 > 0.5$ in 1000G Phase I data, using the appropriate continental subset of 1000G samples (EUR, AMR, or ASN). We used 1000G data, rather than our own exome sequence data, because most reported associations for T2D are with common, intergenic SNVs. We then extended each region of interest by moving out 0.02 cM from those two SNVs (to encompass nearby recombination hotspots), and added an additional 300 kb upstream and downstream. We merged overlapping intervals, yielding 81 unique associated regions, and identified 634 genes completely or partially included within associated regions. In single-variant analyses, we analysed 3,147 non-synonymous variants within these genes in the combined exome sequence and exome array data sets, using a Bonferroni corrected significance threshold of $\alpha = 0.05/3,147 = 1.6 \times 10^{-5}$. We considered gene-level association statistics from exome sequence for these 634 genes using a Bonferroni-corrected significance threshold of $\alpha = 0.05/634 = 7.9 \times 10^{-5}$.

We note that by reducing the stringency of the significance threshold for variants within GWAS loci, we increase the 'experiment-wise' type I error rate across the entire exome. Assuming that 3% of 100,000 coding variants interrogated in

this study map to T2D GWAS loci, as defined above, we would need to change the threshold of significance outside of these regions to $P < 2.1 \times 10^{-8}$ to maintain an 'experiment-wise' type I error rate of 5%.

Testing for 'synthetic associations' at T2D loci in GoT2D genome sequence data.

To identify low-frequency or rare variants that could potentially define synthetic associations, we analysed the ten T2D loci at which a previously-reported tag SNV achieved $P < 0.001$ in our single-variant analysis of the genome sequence data set. We defined as candidates at each locus all low-frequency or rare variants (excluding singletons) within a 5-Mb window (centred on the prior GWAS signals) and tested for synthetic associations caused by either (i) a single low-frequency or rare variant or (ii) multiple low-frequency or rare variants on a common haplotype.

To identify synthetic associations driven by a single low-frequency or rare variant at each of the ten loci, we performed a series of conditional analyses in which we tested for association between gene dosage at the previously reported GWAS index SNV and T2D risk via logistic regression, while including each candidate low-frequency or rare SNV (excluding singletons) as an additional covariate, one-by-one. If inclusion of the low-frequency or rare variant resulted in a conditional association $P > 0.05$ for the tag SNV, we considered the common-variant association signal a potential synthetic association.

To identify synthetic associations based on sets of low-frequency or rare variants, we extended this approach. We (i) defined common haplotypes segregating at each T2D locus; (ii) identified all low-frequency or rare (excluding singletons) variants occurring on T2D-associated haplotypes (haplotypes on which the T2D-associated GWAS index SNV minor allele is present); and (iii) asked whether any combination of these low-frequency or rare variants could explain the effect observed at the T2D GWAS index SNV. We carried out these analyses restricting attention to protein-coding variants within the window and then again for all low-frequency and rare SNVs in the 5-Mb window.

To define common haplotypes at each locus, we used the phased whole-genome sequence data. We first employed the phased genotypes for common (MAF $> 5\%$) variants segregating in the interval between recombination hotspots at the locus (to minimize the number of recombinant haplotypes identified). We next identified the haplotypes on which the T2D-associated (risk or protective) GWAS index SNV minor allele was present. We then assembled the set of low-frequency and rare variants from across the 5-Mb interval which occurred on the background of these T2D-associated common-variant haplotypes. Owing to recombination and imperfect phasing, low-frequency or rare (excluding singletons) variants are often observed on more than one haplotype background. We included all low-frequency or rare variants that occurred more frequently on a T2D-associated haplotype than on other haplotypes.

From this pool of low-frequency and rare variants, we considered only variants with the same direction of effect as the common GWAS index SNV minor allele, as required by the synthetic association hypothesis, which posits that low-frequency or rare variants of larger effect than the common SNV could induce a weaker association signal. We then used a greedy algorithm to select the low-frequency or rare variant which, when added to the index GWAS SNV's dosage in a logistic regression, most reduced the residual effect remaining at the index SNV, as measured by estimated conditional OR. We repeated this process, adding variants to the model, until the estimated effect at the index SNV genotype or gene dosage changed sign, representing no residual effect of the index SNV. At each locus, we also counted the number of variants required to increase the association P value at the GWAS index SNV beyond the nominal $P = 0.05$ significance threshold (Extended Data Table 4).

Credible set analysis of GoT2D genome sequence data. At 78 of the 80 T2D GWAS loci (see *Conditional analyses in established GWAS loci* above), the previously reported index SNV had MAF $> 1\%$ in our GoT2D genome-sequenced sample. At these 78 loci, we constructed credible sets of common variants that, with some minimum specified probability (for example, $\geq 99\%$), contain the variant causal for the corresponding association signal. Our analysis assumes a single causal SNV per signal and that the SNV was genotyped^{30,31}. We constructed credible sets for up to two independent association signals at each locus; at five loci with multiple independent ($r^2 < 0.10$) GWAS index SNVs, we constructed two distinct credible sets.

For each GWAS index SNV, we identified the set of common variants with $r^2 \geq 0.10$ with the index SNV within a 5-Mb window centred on the index SNV. For each variant in this set, we calculated the posterior probability of being causal³¹. We first calculated an approximate Bayes' factor (ABF) for each variant as:

$$ABF = \sqrt{1 - r} e^{rz^2/2}$$

where $r = 0.04/(s.e.^2 + 0.04)$, $z = \beta/s.e.$, and β and s.e. are the estimated effect size (log OR) and its standard error from logistic regression. We then calculated the posterior probability for each variant as ABF/T , where T is the sum of the ABF values over all candidate variants across the interval. This calculation assumes a

Gaussian prior with mean 0 and variance 0.04 for β , the same prior employed in the commonly used single-variant association program SNPTEST⁷².

We based the analysis on the genome-wide meta-analysis results, since most common variants were included in this analysis, and sample sizes were significantly larger than for the genome sequence data alone.

We calculated the effective imputed sample size for each variant in the meta-analysis data as $N_{\text{eff}} = \sum_{j=1}^{13} r_j^2 n_j^{\text{eff}}$, where r_j^2 is the imputation quality and n_j^{eff} is the effective sample size for imputation cohort j . To ensure approximately uniform sample size across variants, we considered to be well-imputed only those variants with effective imputed sample size ($N_{\text{eff}} \geq 80\%$) of the maximum observed across all variants in the window.

Indels were not imputed or meta-analysed in this study, and <2% of common SNVs were not well-imputed by the above effective sample size criterion. To include these common variants while using the most precise estimates available, we calculated posterior probabilities separately from each genome-wide data source. Where an indel from the sequence data set had a SNV proxy in high LD ($r^2 \geq 0.80$) in the meta-analysis data set, we used the proxy's information instead. Where a common SNV that was poorly imputed had high-quality association data from the genome sequence data alone, the posterior probability from the genome sequence data set was used instead. In each case, the final posterior probabilities for all SNVs were re-scaled such that their sum across a locus equaled one.

We used these final posterior probabilities to rank variants in decreasing order. To define credible sets of a specified level (for example, 99%), we included variants with highest final posterior probabilities until their sum reached or exceeded that level (Supplementary Table 28).

Genome enrichment analyses of the GoT2D genome sequence data

Genomic annotation. We collected genome annotation data from several sources. First, we obtained gene transcript information from GENCODEv14 (ref. 80). For protein-coding genes, we included transcripts with a protein-coding tag that either were present in the conserved coding DNA sequence (CCDS) database or had experimentally confirmed mRNA start and end; we then included 5' UTR, exon, and 3' UTR regions from the resulting transcripts. For non-coding genes, we included transcripts with a lncRNA, miRNA, snoRNA, or snRNA tag.

Second, we defined regulatory chromatin states in 12 cell types. We collected sequence reads generated for the following assays: H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K36me3, and CTCF ChIP, in nine ENCODE cell types (GM12878, K562, HepG2, Hsmm, HUVEC, NHEK, NHLF, hESC, HMEC)³², pancreatic islets³⁵, and hASC (adipose stromal cell) pre- and mature adipocytes³³. We mapped reads to hg19 using BWA⁵¹ and used the resulting mapped reads for all cell types to call regulatory states using ChromHMM⁸, assuming ten states. We then assigned names to the resulting state definitions: (i) H3K4me3, H3K27ac (active promoter); (ii) H3K4me3, H3K27ac, H3K4me1 (active enhancer 1); (iii) H3K27ac, H3K4me1 (active enhancer 2); (iv) H3K4me1 (weak enhancer); (v) H3K27me3, H3K4me3, H3K4me1 (poised promoter); (vi) H3K27me3 (repressed); (vii) low/no signal 1; (viii) CTCF (insulator); (ix) low/no signal 2; and (x) H3K36me3 (transcription).

Third, we obtained transcription factor-binding ChIP sites from three sources: 141 proteins from ENCODE³², 5 from Pasquali *et al.*³⁵, and 1 from Mikkelsen *et al.*³³.

From gene transcript data we defined CDS (protein coding transcript exons); ncRNA (non-coding RNA transcripts); and 3' and 5' UTR (UTR regions of coding transcripts). From chromatin state data for each of the 12 cell types we identified active enhancers (pooled active enhancer 1 and 2 elements); weak enhancers; and active promoters. From transcription factor binding sites we defined transcription factor-binding sites (TFBS) (sites pooled across all factors). This resulted in a total of 41 annotation categories (Extended Data Fig. 6).

Enrichment of genome annotation. We jointly modelled variants in credible sets using T2D association and the functional annotation classes using the method described by Pickrell³⁸. First, we tested each annotation individually and identified the annotation that most improved the model likelihood. We then iteratively added annotations in this manner until the likelihood did not increase further. Using this set of annotations, we tested a range of penalized likelihoods (from 0 to 1 in 0.01 increments) using tenfold cross-validation, and identified the penalty that gave the best cross-validation likelihood. Using this penalty, we then iteratively dropped annotations to identify the model with the maximal cross-validation likelihood. The resulting model included coding exons, TFBS, hASC mature adipose active enhancers and promoters, pancreatic islet active and weak enhancers and active promoters, hASC pre-adipose active and weak enhancers, NHEK active enhancers, NHLF active enhancers, K562 weak enhancers, HMEC weak enhancers and active promoters, H1-hESC active promoters, ncRNA, and 5' and 3' UTR (Extended Data Fig. 6). Finally, we used this model to update posterior probabilities for each variant and re-calculate 99% credible sets.

Gene enrichment analyses in the GoT2D+T2D-GENES exome sequence data.

We first used the SMP (statistics/matrix/permutation) gene-set enrichment procedure implemented in the PLINK/Seq package (<https://atgu.mgh.harvard.edu/plinkseq/>). This approach calculates enrichment statistics for large sets of genes to establish whether case-enrichment of rare variants is preferentially concentrated in a particular set of genes, controlling for any exome-wide/baseline difference in case and control rates. The procedure uses gene-based association statistics, and forms sums of these statistics over all genes in a set, the significance of which is evaluated by permutation. We considered the relative enrichment statistic, SSET/SEXOME, with significance evaluated empirically (10,000 replicates) based on the null distribution of this ratio. The reported effect sizes from the gene-set enrichment analysis are estimates of the unconditional odds ratio that do not take exome-wide differences in case/control rates into account⁷⁰. We selected 18 'premium' sets of genes (Supplementary Table 32) that reflect the current knowledge of pathways ($N = 15$) involved in T2D and the three sets of genes involved in monogenic form of diabetes defined above: 'Monogenic all' ($N = 81$); 'Monogenic primary' ($N = 28$); and 'Monogenic OMIM' ($N = 13$). We restricted these analyses to singleton and ultra-rare ($\text{MAF} < 0.1\%$) protein-truncating variants.

We then used biological knowledge to test for enrichment of association signal across established sets of genes from Gene Ontology, KEGG, Reactome, and Biocarta collections from MSigDB (version 4.0) as well as a number of hand-curated gene-sets (Supplementary Table 32) that had been generated for the SMP analyses. These analyses calculated measures of gene-set enrichment from gene-level association results (that is, from SKAT-O) by means of a pre-ranked GSEA⁸² method (version 2.0.13), which consists of a weighted Kolmogorov-Smirnov (random bridge) statistic. In our analysis we performed 10,000 permutations on gene-set sizes from 5 to 5,000 genes.

Investigation of genes implicated in Mendelian forms of diabetes in the exome data.

We first curated a list of 81 genes termed the 'Monogenic all' gene-set (Supplementary Table 22), consisting of genes with pathogenic mutations reported to co-segregate with diabetes or a syndrome associated with an increased prevalence of diabetes. Two subsets of the 'Monogenic all' gene-set were then additionally defined: the 'Monogenic primary' gene-set ($N = 28$), consisting of genes with mutations leading to diabetes as a primary feature, and the 'Monogenic OMIM' gene-set ($N = 13$), consisting of genes linked to MODY or neonatal diabetes in the OMIM catalogue (entries 606391 and 606176). In addition to examining the significance of single-variant and gene-based tests within these gene-sets, we also performed an aggregate analysis of all variants in the gene-set. For each of the three gene-sets, we constructed five variant lists by applying the same four masks as in the exome-wide gene-level analysis (PTV-only, PTV + missense, PTV + NS_{broad} and PTV + NS_{strict}), as well as an additional mask containing all variants reported as 'high confidence' and 'disease-causing' in the Human Gene Mutation Database (HGMD), annotated using Biobase 'GenomeTrax' software (<http://www.biobase-international.com/product/genome-trax>). We then analysed each of the fifteen variant lists with the SKAT-O test, using the same meta-analysis procedure and covariates as in the exome-wide gene-based analysis. To obtain effect-size estimates, for each variant list we applied a collapsing burden test, in which logistic regression of T2D status was performed on individual genotypes encoded as 0 (if they carried no variants in the list) or 1 (if they carried at least one variant in the list). Effect size estimates and standard errors were determined using the Firth penalized likelihood method. Analysis in the exome array data set was performed by first generating fifteen variant lists based on the content of the exome array, computing the collapsing burden test for each cohort, and then combining associations and effect size estimates using an inverse variance weighted meta-analysis. To compare the age of diagnosis of variant carriers to those of non-carriers, we used a two-sided t -test.

Protein-protein interaction analyses in the exome data. We performed data-driven extraction of association signal enriched sub-networks (rather than relying on pre-defined gene-sets) from protein-protein interaction (PPI) data. We used two different approaches, both run using the curated PPI database InWeb3 (ref. 83).

The first approach consists of two steps. First, the entire human PPI network was searched for protein complexes (clusters) using the algorithm implemented in clusterONE⁸⁴, which identifies protein complexes with high cohesiveness. The method was run with default parameter settings (0.3 as density threshold, 0.8 as merging threshold, and 2 as the penalty-value node), and with the-fluff option activated, which allows the addition of highly connected boundary nodes to the cluster. Second, gene-based association P values derived from SKAT-O analyses of the 12,940 multiethnic exome sequences were aggregated, using Fisher's method, for the genes encoding each of the proteins within a cluster to generate a 'cluster association' statistic.

An empirical P value for the significance of these aggregated cluster association statistics was derived by comparing each cluster to a large number of complexes of the same topology, but composed of randomly sampled proteins. Specifically,

a background distribution was obtained for each protein complex as follows: each protein in the cluster was randomly substituted by a different protein represented in the InWeb3 database, matched for number of minor allele carriers in the data set. SKAT-O P values were assigned to each protein from the exome sequencing results, and an aggregated P value was obtained for each pseudo-complex using Fisher's method, as above. This process was repeated 100,000 times, and the empirical P value for each complex was calculated as the proportion of the iterations for which the Fisher's P value of the observed complex was more significant than that of P values for the pseudo-complexes. This procedure was repeated for all gene-level masks (PTV-only, PTV + missense, PTV + NS_{strict} and PTV + NS_{broad}).

To test the study-wide significance of apparently associated clusters, we used two permutation designs. In the first design, we generated 100,000 pseudo-complexes for each cluster, replacing each protein within each cluster with one protein from InWeb3, matched for the number of minor allele carriers in the data set. We calculated the number of permuted data sets which generated any 'pseudo-cluster' association P value more significant than our most enriched cluster. In the second design, we used a Monte-Carlo algorithm to generate 10,000 random PPI networks, with the same degree as observed in the InWeb3 database, ran clusterONE on each, and once again compared the distribution of 'best' cluster association P value with that observed in the real data.

The second approach uses the dense module searching algorithm (a heuristic 'greedy' method) described in dmGWAS⁸⁵, where a module is defined as a sub-network within the whole network if it contains a locally increased proportion of low P value genes. This method differs from the earlier method in using the association P values, in combination with the PPI data, to construct the networks. The module is grown for each protein in the PPI by adding the neighbouring nodes within a pre-defined distance ($d=2$) that can yield a maximum increment of the module score $Z_m^{(k)} = \sum Z_i / \sqrt{k}$ for module m , where k is the number of genes in the module and Z_i is calculated from the P value of exome gene-based tests using an inverse normal distribution function. The addition of neighbourhood nodes is stopped when the increment is less than 10% of $Z_m^{(k)}$ (that is, $Z_m^{(k+1)} < Z_m^{(k)} + (Z_m^{(k)} \times 0.1)$). As with the clusterONE approach, this procedure was conducted for all four exome gene-based level masks.

To evaluate whether the top ranked-modules are significantly associated with T2D, we permuted case-control status across the 12,940 exomes (maintaining ethnic strata) 10,000 times and generated 10,000 SKAT-O gene-based association tests on all genes in the top 15 modules (once for each gene-based variant mask, 40,000 in total). During each permutation, Z_m was re-calculated for each module, and a set of empirical P values was obtained by comparing the P value of the original module to these modules with the SKAT-O results from the swapped labels. Following the above procedure, all 15 top modules were found significantly enriched for the PTV + NS_{strict} and PTV + NS_{broad} gene-based variant masks ($P < 10^{-4}$, after the 10,000 case-control permutations).

Modelling disease architecture

T2D liability risk and architecture bounding in the exome array data. We used a Bayesian framework implemented in R to compute the probability that each variant explains more than a defined amount of the T2D-risk liability-scale variance (LVE). The joint distribution in the MAF-OR space is computed by assuming a T2D prevalence of 8% and beta and normal distributions for the MAF and OR, respectively. The OR is calculated with reference to the minor allele. The MAF is adjusted to take account of apparent allele frequency heterogeneity between cohorts (subjects from missing cohorts are excluded from calculations). Analyses are restricted to variants with MAF > 0.1% because the representation of variants with MAF below this threshold on the exome array is poor. The probability is obtained by numerically integrating over the joint distribution for MAF-OR combinations that explain more than the defined amount of liability-scale variance. For bounding the maximum number of variants that could contribute to T2D risk variance, we performed a sensitivity analysis on the 88 known T2D index SNVs present on the exome array to define the thresholded variance explained and the probability: this analysis shows that for a probability of >0.8 to explain 0.01% of the T2D risk variance, we were able to identify 91% of these known T2D SNVs. Ranges of OR and MAF consistent with 80% power to detect single-variant association in this data set (for exome-wide significance, $P < 5 \times 10^{-7}$) were calculated to reflect the fact that differences in sample size for individual variants (due to differences in allele frequency distribution and genotyping QC) also influence power. The relationship between power and LVE differs for risk and protective alleles because of unequal numbers of cases and controls.

Genetic architecture simulations based on GoT2D data and results

Range of simulated disease models. Following our previously published framework⁴⁰, we conducted population genetic simulations of T2D architecture using the forward simulation program ForSim⁸⁶. We assumed T2D prevalence 8% and heritability ~45%, and chose the mutation rate, recombination rate, a gamma

distribution of selection coefficients, and other parameters of demographic history by fitting the simulated site frequency spectrum to empirical high-coverage exome sequence data from GoT2D.

We then considered a wide range of disease models by varying two parameters: coupling parameter τ , which regulates how strongly selection against a disease-causing allele depends on the per-allele disease risk⁸⁷; and target size T , the summed lengths of the genomic regions within which mutations can influence T2D risk. Specifically, a variant's additive contribution to disease risk g is given by $g = s\tau(1 + \varepsilon)$ where s is the selection coefficient under which the variant evolves and ε is drawn from a normal distribution⁴⁰.

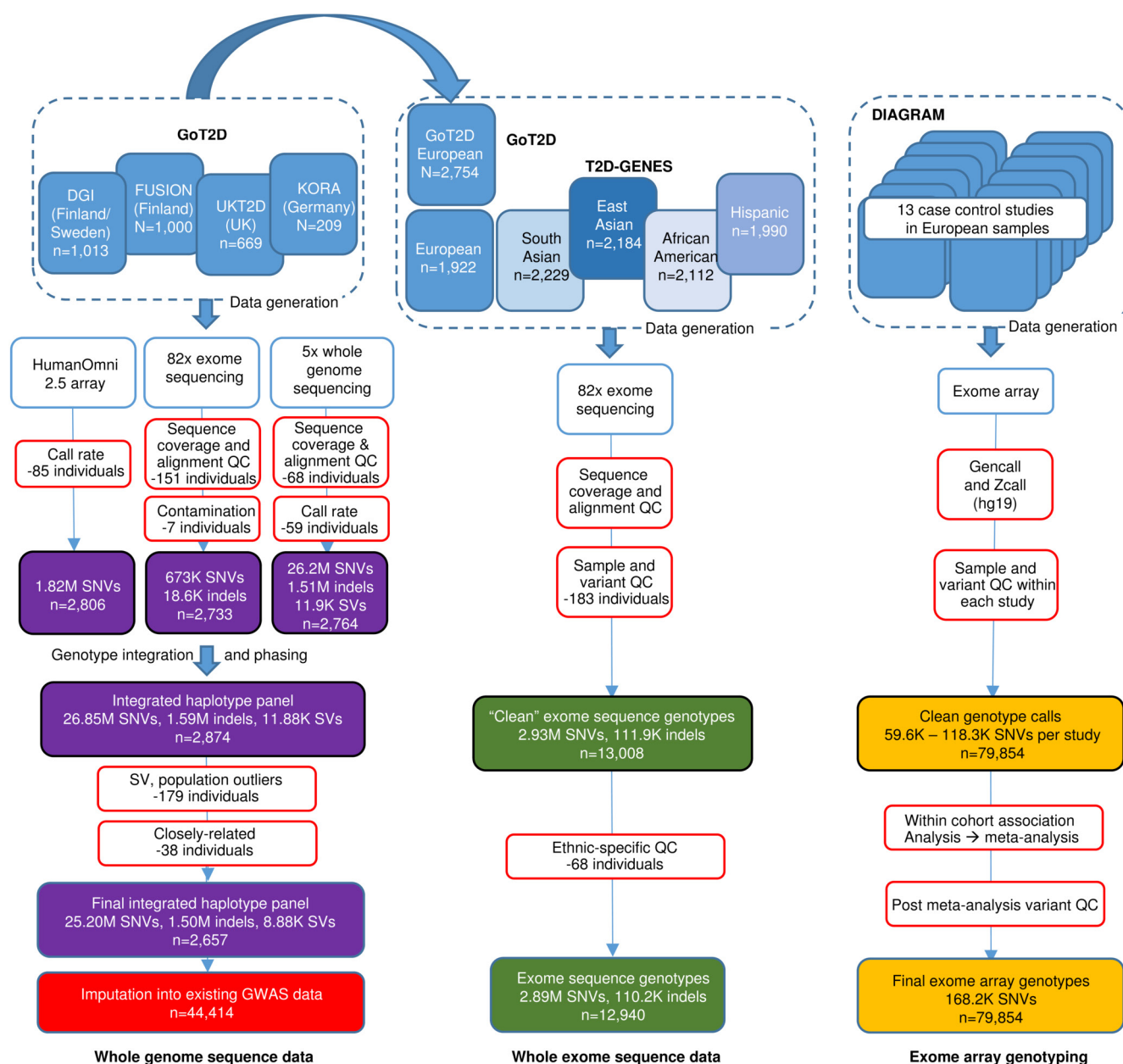
By varying τ and T , we generated a wide range of joint distributions for allele frequency and effect size. In total, we evaluated 12 models: $\tau = 0, 0.1, 0.3$, and 0.5 crossed with $T = 750$ kb, 2.0 Mb, and 3.75 Mb. Under models with higher selection against strongly deleterious alleles (larger τ), rare variants explain the bulk of heritability and can have large effects, while under models with weak dependence (smaller τ), common variants explain the bulk of heritability and rare variants collectively have weaker effects. Although we had previously excluded many models as producing predictions inconsistent with observed sibling relative risk, GWAS, and linkage results, prior work showed that models varying widely in the proportion of total heritability attributable to rare versus common variation were still plausible⁸⁸. In this study, we explored whether the space of plausible disease models could be further constrained using whole genome sequence, imputation, and meta-analysis results.

Simulation procedure ForSim enables simulation of variants across user-specified loci in large populations. Inputs include a demographic history (trained on European sequence data) and a gamma distribution of selection coefficients for a subset of variants under natural selection. We simulated genotypes for a current population of effective size 500,000 individuals⁴⁰ and selected potential disease risk variants from those under selection appropriate to the intended target size. Each risk variant received a disease-specific effect size depending on the selection coefficient under which it evolved and the assumed degree of dependence between selection and effect size. Each individual was then designated as case or control depending on his/her cumulative genetic risk score plus a random environmental risk component chosen to achieve the estimated T2D heritability of ~45%. From this population simulated with both phenotypes and genotypes, we selected appropriate numbers of cases and controls and conducted single-variant association tests in order to compare the distribution of P values from simulation to that observed in the current study. Results shown are the average of 25 independent simulation replicates for each disease model.

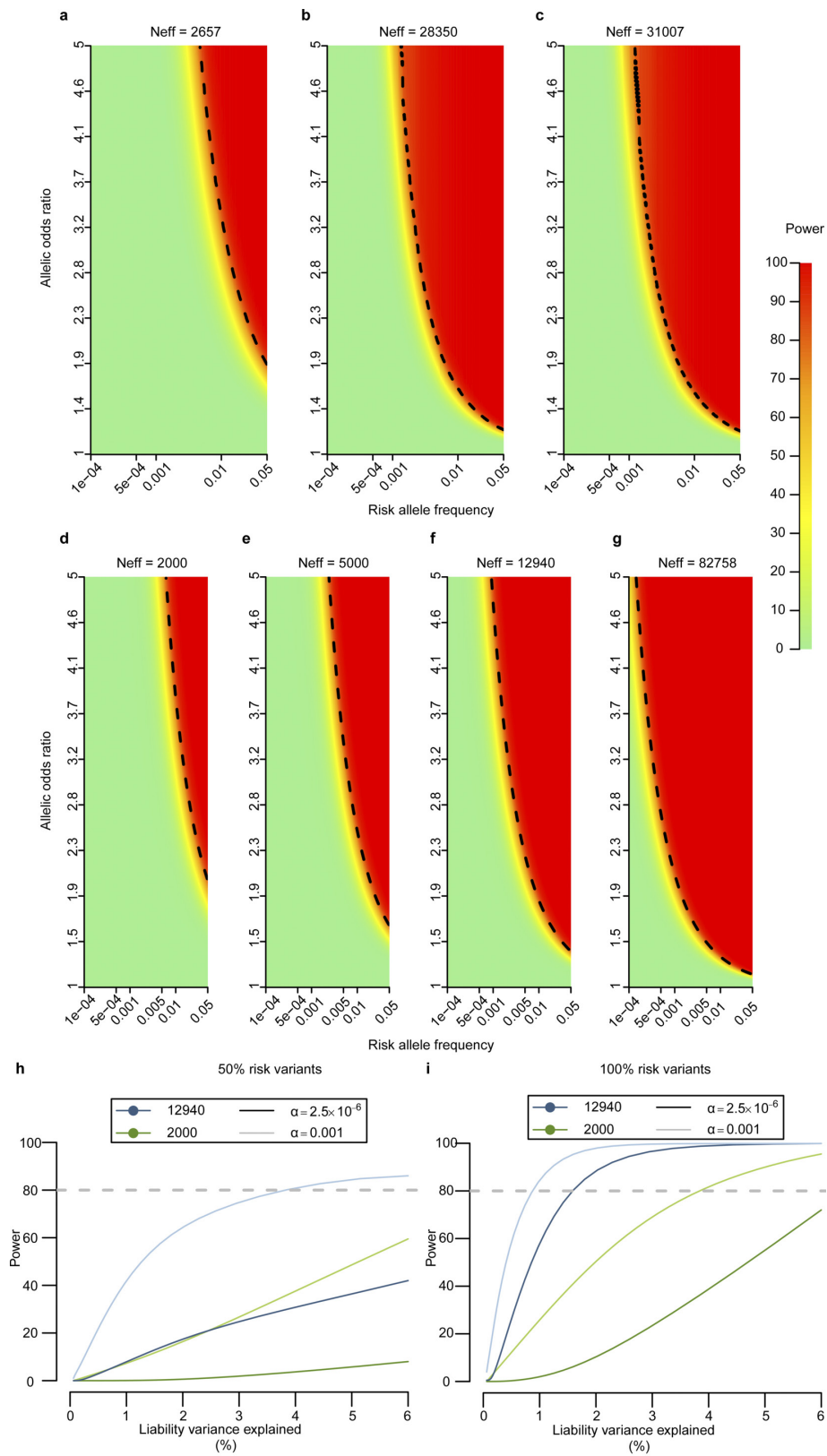
Comparison of simulated outcomes to empirical T2D results. We focused on comparing simulated outcomes under three disease models, each of which were previously found to be consistent with sibling relative risk, GWAS, and linkage results for T2D, but vary widely in causal variant properties (Fig. 3): a rare-variant model in which rare variants explain ~75% of T2D heritability (small target size $T = 750$ kb and moderate dependence between effect size and selection $\tau = 0.5$), an intermediate model in which rare, low-frequency, and common variants all contribute significantly to T2D heritability ($T = 2.0$ Mb and $\tau = 0.3$), and a common polygenic model in which common variants explain ~75% of T2D heritability ($T = 3.75$ Mb and weak dependence $\tau = 0.1$). We first compared the simulated outcomes of a whole-genome sequencing study in ~3,000 samples under each model. All three models predicted similar distributions of variant association test statistics using the sequenced individuals alone (data not shown).

However, the predictions began to diverge when we simulated imputation into GWAS samples and studied the distribution of test statistics after meta-analysis. For each simulated model, we sampled 14,175 cases and 14,175 controls (to match the effective sample size of the actual imputation cohorts used for meta-analysis). Because genotyping accuracy in simulated samples is perfect (unlike in imputation), we calculated average imputation quality as a function of MAC in the empirical data (using the r^2 value reported by the imputation software that was used in each cohort). We then corrected, for each variant, the association test statistic in simulated data by multiplying the χ^2 value by the average imputation r^2 for the variant MAC. We then re-computed association P values from the corrected chi-squared statistics to compare P value distributions in simulated versus empirical data. We plotted the distribution of association P values for variants of different frequency classes in a quantile-quantile (QQ) plot, and compared these curves to the empirical T2D results (Fig. 3). Focusing on low-frequency variants, we also asked how many unique low-frequency signals achieved significant association to T2D risk under each simulated model, and compared these quantities to empirical observation (Fig. 3). These analyses demonstrate that the intermediate and rare-variant models produce an excess of association signal among low-frequency variants compared to observation, whereas the common polygenic model is consistent with the genome-wide distribution of association signals observed.

50. Guey, L. T. *et al.* Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genet. Epidemiol.* **35**, 236–246 (2011).
51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
52. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
53. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
54. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).
55. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
56. Handsaker, R. E., Korn, J. M., Nemesh, J. & McCarroll, S. A. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat. Genet.* **43**, 269–276 (2011).
57. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
58. Li, Y., Sidore, C., Kang, H. M., Boehnke, M. & Abecasis, G. R. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res.* **21**, 940–951 (2011).
59. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135, author reply 135–139 (2008).
60. Weale, M. E. Quality control for genome-wide association studies. *Methods Mol. Biol.* **628**, 341–372 (2010).
61. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
62. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
63. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
64. Firth, D. Bias reduction of maximum-likelihood-estimates. *Biometrika* **80**, 27–38 (1993).
65. Ma, C., Blackwell, T., Boehnke, M. & Scott, L. J. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genet. Epidemiol.* **37**, 539–550 (2013).
66. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
67. Seldin, M. F., Pasienc, B. & Price, A. L. New approaches to disease mapping in admixed populations. *Nat. Rev. Genet.* **12**, 523–528 (2011).
68. Price, A. L. *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* **5**, e1000519 (2009).
69. Churchhouse, C. & Marchini, J. Multiway admixture deconvolution using phased or unphased ancestral panels. *Genet. Epidemiol.* **37**, 1–12 (2013).
70. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
71. Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762–775 (2012).
72. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
73. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
74. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
75. Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
76. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).
77. Rice, W. R. A consensus combined *P*-value test and the family-wide significance of component tests. *Biometrics* **46**, 303–308 (1990).
78. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
79. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
80. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
81. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
82. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
83. Lage, K. *et al.* A human phenotype-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–316 (2007).
84. Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein–protein interaction networks. *Nat. Methods* **9**, 471–472 (2012).
85. Jia, P., Zheng, S., Long, J., Zheng, W. & Zhao, Z. dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks. *Bioinformatics* **27**, 95–102 (2011).
86. Lambert, B. W., Terwilliger, J. D. & Weiss, K. M. ForSim: a tool for exploring the genetic architecture of complex traits with controlled truth. *Bioinformatics* **24**, 1821–1822 (2008).
87. Eyre-Walker, A. Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl Acad. Sci. USA* **107** (Suppl 1), 1752–1756 (2010).
88. Lyssenko, V. *et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N. Engl. J. Med.* **359**, 2220–2232 (2008).



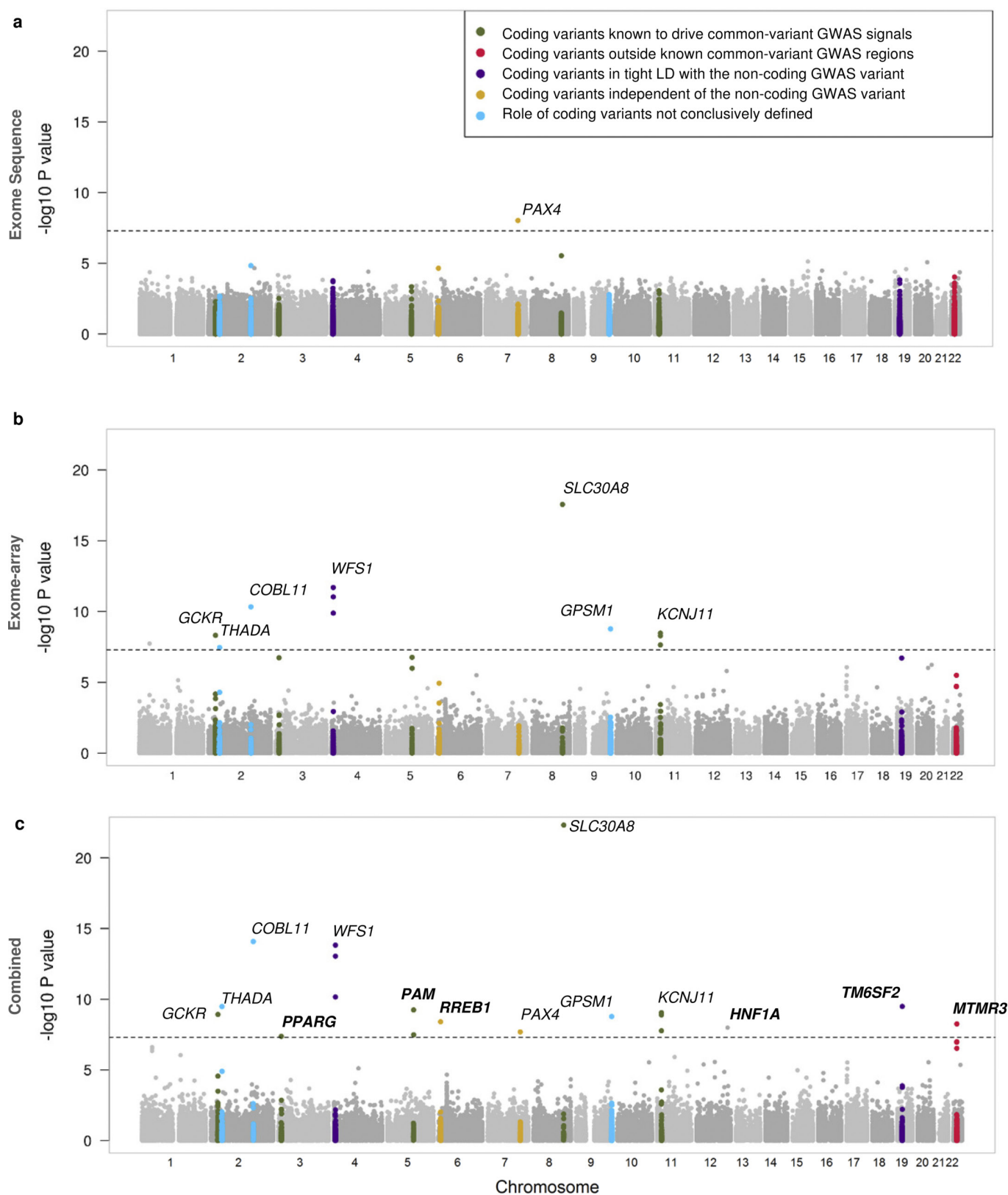
Extended Data Figure 1 | Summary of samples and quality control procedures. This figure summarizes data generation for whole-genome sequencing (GoT2D), exome sequencing (GoT2D and T2D-GENES), exome array genotyping (DIAGRAM), and GWAS imputation (DIAGRAM).



Extended Data Figure 2 | See next page for caption.

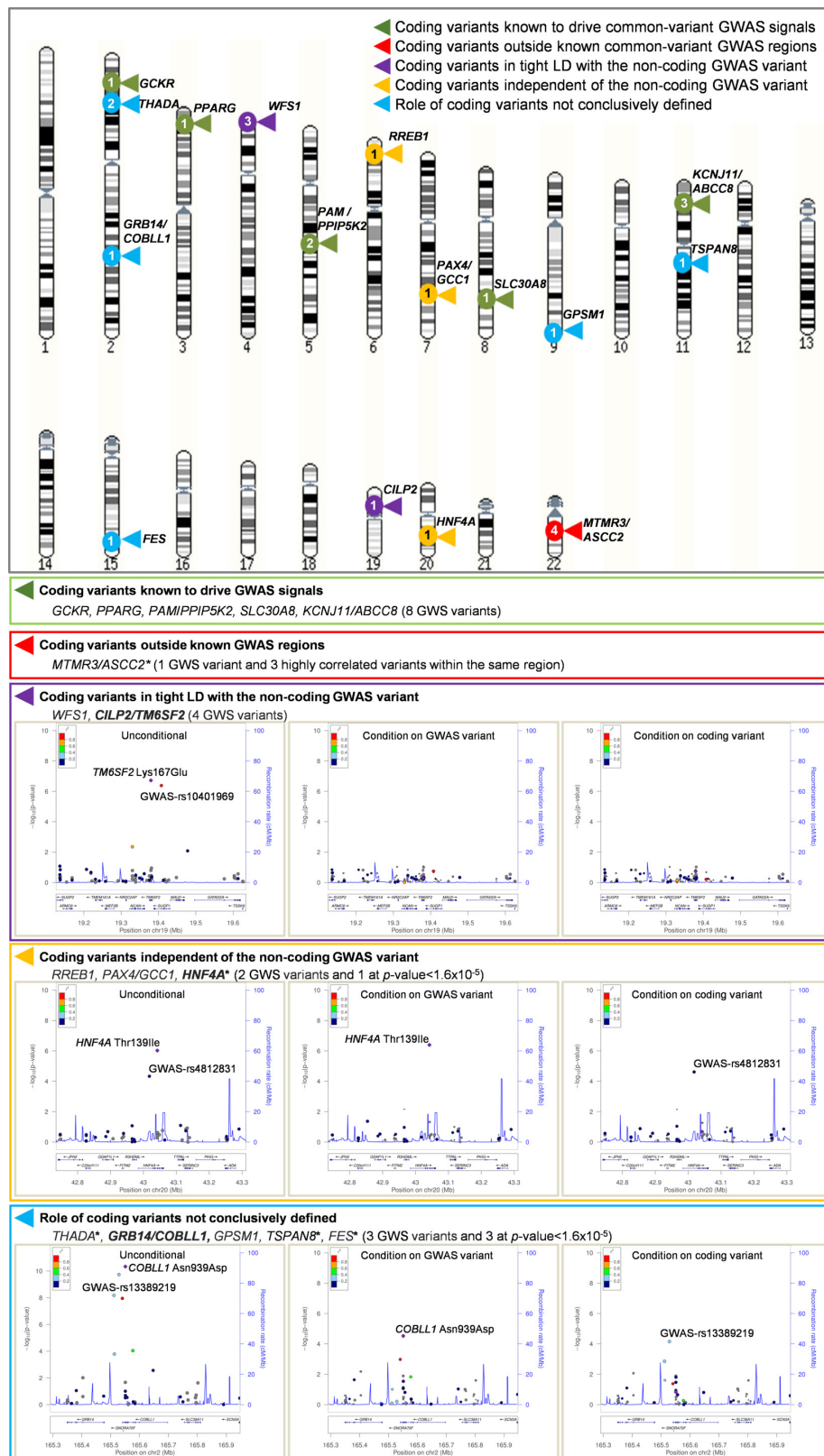
Extended Data Figure 2 | Power for single and aggregate variant association. **a–g**, Power to detect single-variant association ($\alpha = 5 \times 10^{-8}$) at varying minor allele frequencies (x -axis) and allelic ORs (y -axis) for seven effective sample size (N_{eff}) scenarios relevant to the genomes (**a–c**) and exomes (**d–g**) components of this project. **a**, Variant observed in 2,657 samples (the effective size of the GoT2D integrated panel). **b**, Variant observed in 28,350 samples (the effective size of the imputed data set). **c**, Variant observed in the GoT2D integrated panel and the imputed data set (effective sample size 31,007). **d**, Ancestry-specific variant in 2,000 samples (the size of each of the non-European exome sequence data sets). **e**, European-specific variant in 5,000 samples (the combined size of the European exome sequence data sets). **f**, Variant observed with shared frequency across all ancestry groups in 12,940 samples (the size of the

combined exome sequence data set). **g**, Variant observed in the combined exome array and sequencing data set (effective sample size 82,758). **h, i**, Power for gene-based test of association (SKAT-O) according to liability variance explained. In **h**, 50% of the variants contribute to disease risk and the remaining 50% have no effect on disease risk; in **i**, 100% of the variants contribute to disease risk. For each, sample sizes considered are 2,000 (ancestry-specific effects; green) and 12,940 (ancestry-shared effects; blue). Power is shown for two levels of significance ($\alpha = 2.5 \times 10^{-6}$ and $\alpha = 0.001$). From these simulation studies, it is clear that under the optimistic model, where effects are shared across all ethnicities (blue line) and all variants contribute, power is >60% for 1% variance explained and $\alpha = 2.5 \times 10^{-6}$. However, power declines rapidly if either criterion is relaxed.



Extended Data Figure 3 | Single variant analyses. a–c, Manhattan plot of single-variant analyses generated from exome sequence data in 6,504 cases and 6,436 controls of African American, East Asian, European, Hispanic, and South Asian ancestry (a); exome array genotypes in 28,305 cases and 51,549 controls of European ancestry (b); and combined meta-analysis of exome array and exome sequence samples (c). Coding variants are categorized according to their relationships to the previously

reported lead variant from GWAS region. Loci achieving genome-wide significance only in the combined analysis are highlighted in bold. The *HNF1A* variant reaching genome-wide significance in the combined analysis is a synonymous variant (Thr515Thr). The dashed horizontal line in each panel designates the threshold for genome-wide significance ($P < 5 \times 10^{-8}$).

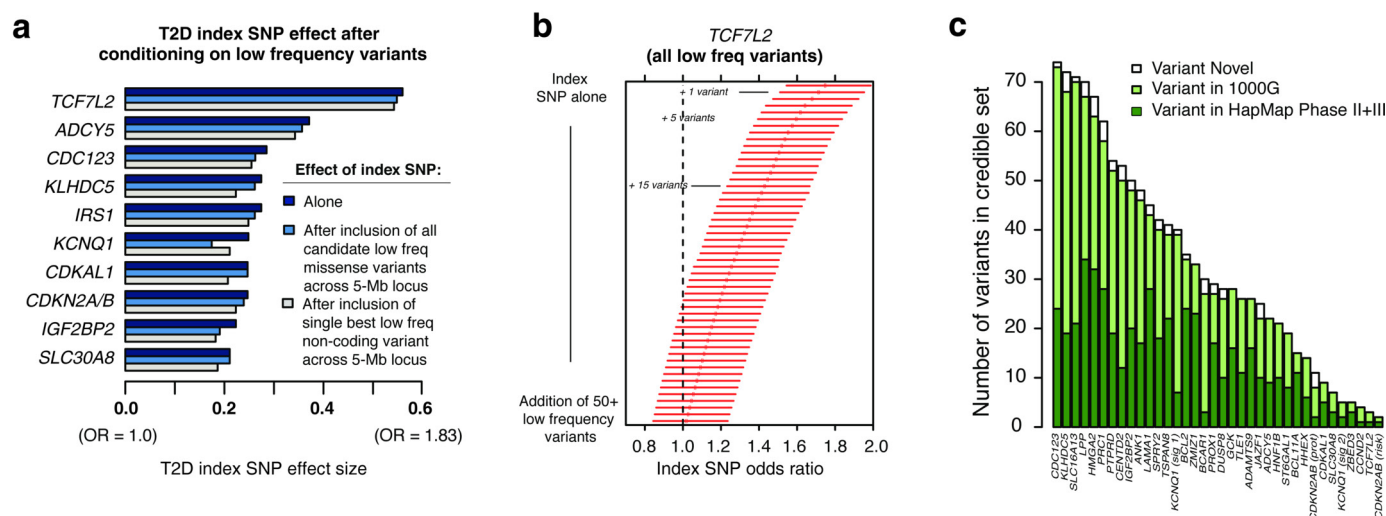


Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Classification of coding variants according to their relationship to reported lead variants for each GWAS region.

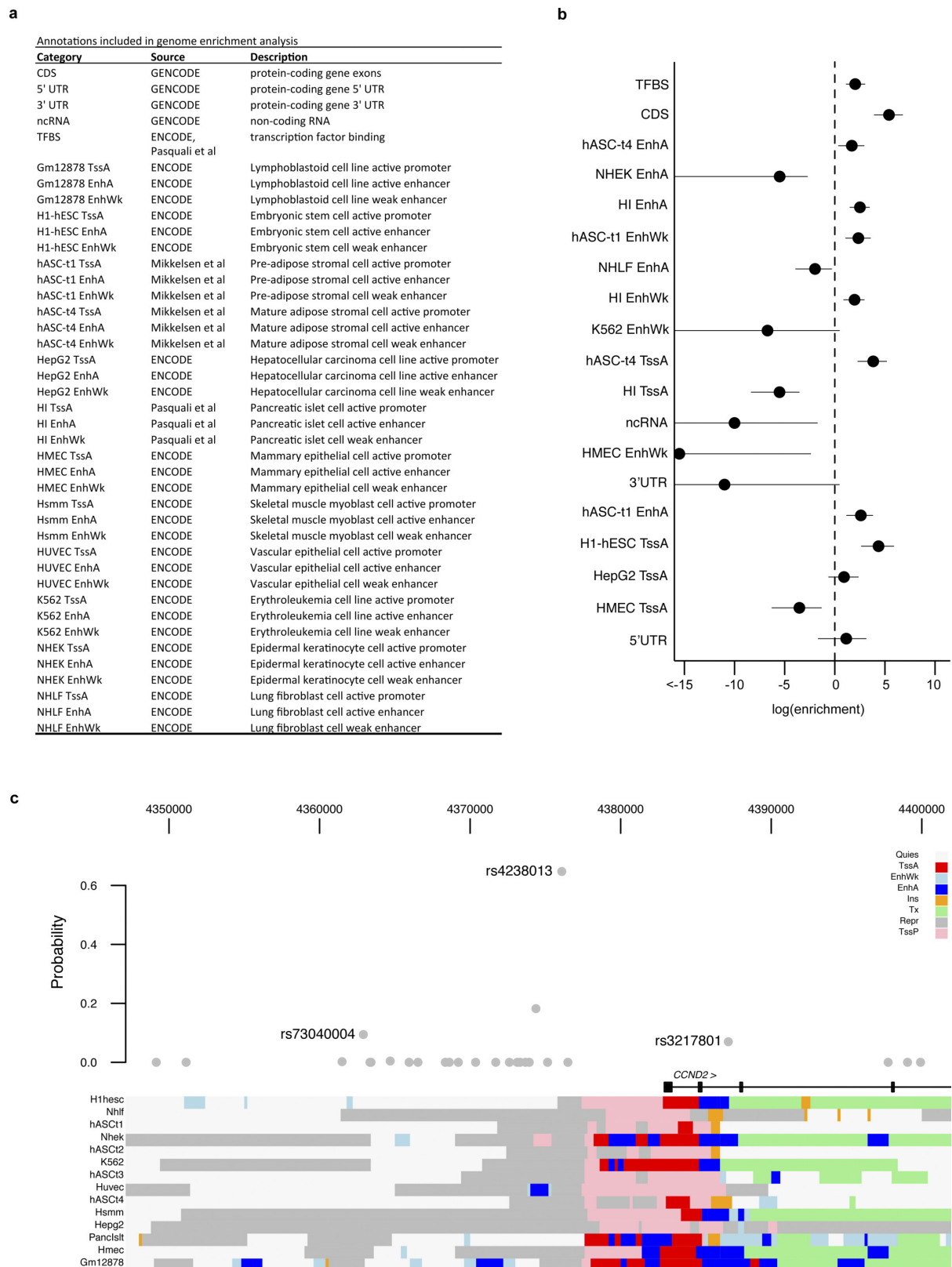
The ideogram shows the location of 25 coding variant associations at 16 loci described in the text. The number in each circle corresponds to the number of associated variants at each locus. Variants are grouped into five categories based on inferred relationship with the GWAS lead variant. For some of these categories, the figure includes representative regional association plots based on exome array meta-analysis data from 28,305 cases and 51,549 controls. The locus displayed for each category is designated in bold. The first plot in each panel shows the unconditional association results; the middle plot the association results after conditioning on the non-coding GWAS SNP; and the last plot the results after conditioning on the most significantly associated coding variant.

Each point represents an SNP in the exome array meta-analysis, plotted with its P value (on a $-\log_{10}$ scale) as a function of the genomic position (hg19). In each panel, the lead coding variant is represented by the purple symbol. The colour-coding of all other SNPs indicates LD with the lead SNP (estimated by European r^2 from 1000G March 2012 reference panel: red $r^2 \geq 0.8$; gold $0.6 \leq r^2 < 0.8$; green $0.4 \leq r^2 < 0.6$; cyan $0.2 \leq r^2 < 0.4$; blue $r^2 < 0.2$; grey r^2 unknown). Gene annotations are taken from the University of California Santa Cruz genome browser. GWS: genome-wide significance. *Seven variants, three at *ASCC2*, and one each at *THADA*, *TSPAN8*, *FES* and *HNF4A* did not achieve genome-wide significance themselves, but are included because they fall into genes and/or regions with other significant association signals (see text).



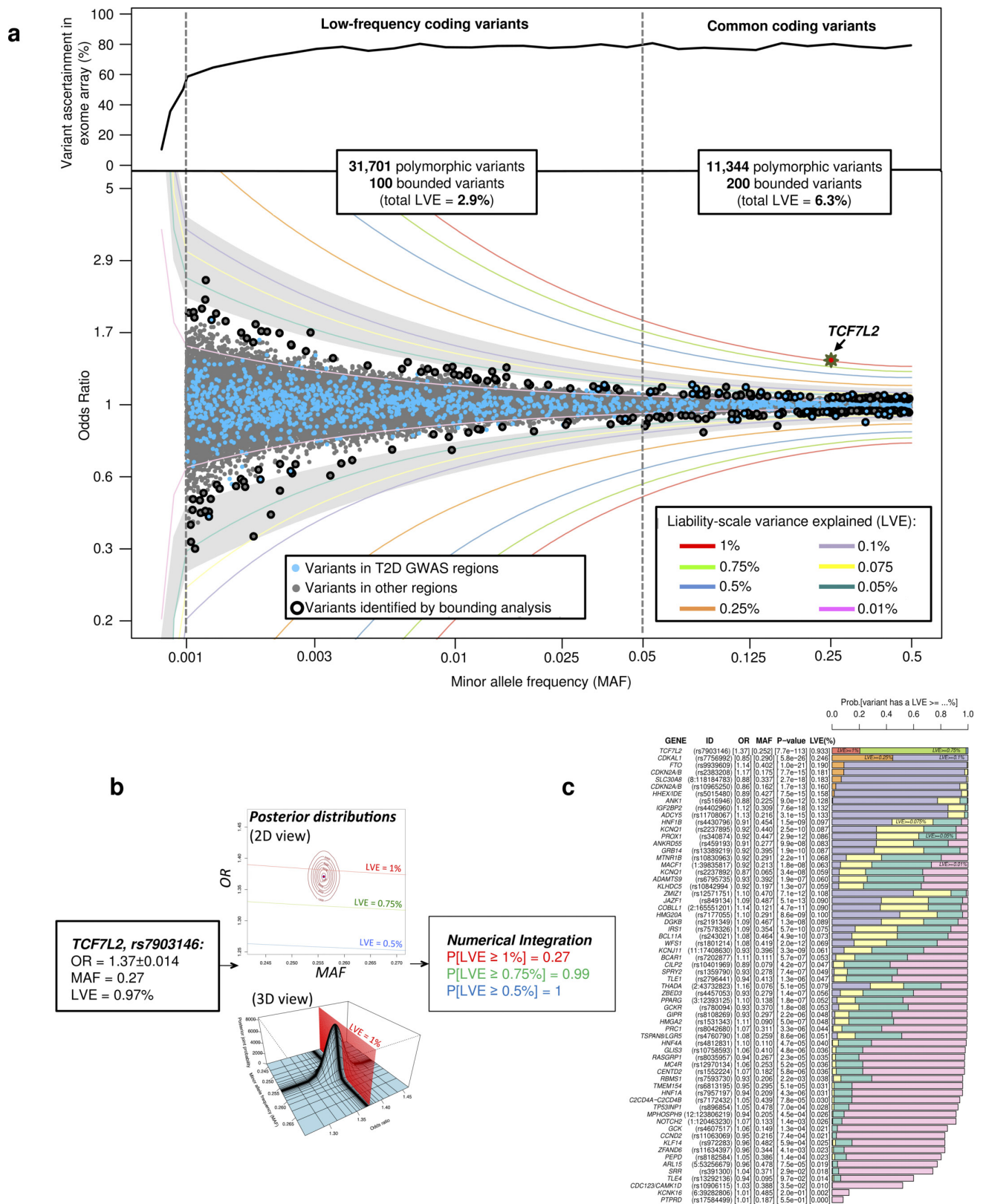
Extended Data Figure 5 | Exclusion of synthetic associations and construction of credible causal variant sets at T2D GWAS loci. Ten T2D GWAS loci were selected for synthetic association testing ($P < 0.001$; see Methods). **a**, The effect size observed at the GWAS index SNV (sequence data) before (navy blue) and after (light blue, grey) conditioning on candidate rare and low-frequency (MAF $< 5\%$) variants which could

produce synthetic association. **b**, Example of synthetic association exclusion at the *TCF7L2* locus. Error bars represent 95% confidence intervals for the index SNP odds ratio as rare variants are greedily added to the model. **c**, The size of credible sets at T2D GWAS loci when constructed from the GoT2D data, compared to the sizes when restricted to variants in the 1000G or HapMap data.



Extended Data Figure 6 | Genome enrichment analysis in GoT2D whole genome sequence data. $n = 2,657$. **a**, Functional annotation categories were defined using transcription, chromatin state and transcription factor binding data from GENCODE, ENCODE and other studies. **b**, T2D association statistics for variants at each T2D locus were jointly modelled with functional annotation using fgwas. In the resulting model we identified enrichment of coding exons (CDS), transcription factor binding sites (TFBS), mature adipose active enhancers and promoters

(hASC-t4 EnhA, TssA), pancreatic islet active and weak enhancers (HI EnhA, EnhWk), pre-adipose active and weak enhancers (hASC-t1 EnhA, EnhWk), embryonic stem cell active promoters (H1-hESC TssA) and 5'UTRs. Dots represent enrichment estimates and horizontal lines the 95% confidence intervals. **c**, At the *CCND2* locus, three variants not present in HapMap2 have a combined 90% posterior probability of being causal (rs4238013, rs3217801, rs73040004). One of these variants, rs3217801, is a 2-bp indel that overlaps an islet enhancer element.



Extended Data Figure 7 | Low frequency variants in exome array data.

Results from meta-analysis of 43,045 low-frequency and common coding variants on the exome array (assayed in 79,854 European subjects).

a, Observed allelic ORs as a property of allele MAF. Variants missing in more than eight cohorts or polymorphic in only one cohort were excluded. Coloured lines represent contours for liability variance explained. Regions shaded grey denote ranges of OR and MAF consistent with 80% power (in this case, at $\alpha = 5 \times 10^{-7}$) to detect single-variant associations in this

data set (given the observed range of missing data). Variants with a black collar are those highlighted by a bounding analysis as having a probability > 0.8 of having liability-scale variance (LVE) $> 0.1\%$. **b**, Distribution of each variant in the MAF/OR space was computed by assuming T2D prevalence of 8% and a beta and normal distribution for MAF and OR, respectively. Probability is obtained by integrating the joint MAF-OR distributions over ranges of LVE. **c**, Single variant association, liability and bounding results for the known T2D GWAS variants on the exome array (see Methods).

Extended Data Table 1 | Summary information for sample sets used in the association analyses

Ancestry	Study	Countries of Origin	Num. of Cases (% female)	Num. of Controls (% female)	Effective Sample Size
Whole Genome Sequencing Studies					
European	Finland-United States Investigation of NIDDM Genetics (FUSION) Study	Finland	493 (41.5)	486 (45.2)	979
European	Kooperative Gesundheitsforschung in der Region Augsburg (KORA)	Germany	101 (44.5)	104 (66.3)	205
European	Malmo-Botnia Study	Finland, Sweden	410 (51.5)	419 (44.1)	829
European	UK Type 2 Diabetes Genetics Consortium (UKT2D)	UK	322 (46.2)	322 (82.2)	644
Total Whole Genome Sequence			1,326	1,331	2,657
Genome-Wide Array Studies					
European	INTERACT	France, Germany, Italy, Netherlands, Spain, Sweden, UK	4624 (51.8)	4668 (64.2)	9292
European	Wellcome Trust Case Control Consortium (WTCCC)	UK	1586 (40.9)	2938 (50.8)	4120
European	Kooperative Gesundheitsforschung in der Region Augsburg (KORA)	Germany	993 (45.1)	2985 (52.2)	2980
European	Framingham Heart Study (FHS)	US	673 (42.6)	7660 (55.1)	2475
European	Finland-United States Investigation of NIDDM Genetics (FUSION) Study	Finland	1060 (43.1)	1090 (51.3)	2150
European	Diabetes Genetics Initiative (DGI)	Finland, Sweden	899 (46.6)	1057 (49.6)	1943
European	Estonian Genome Center, University of Tartu (EGCUT-OMNI)	Estonia	389 (58.6)	6013 (54.2)	1461
European	Diabetes Gene Discovery Group (DGDG)	France, Canada	677 (39.3)	697 (59.7)	1374
European	Mt Sinai BioMe Biobank Platform (BioMe (Illumina))	US	255 (29.0)	1647 (51.4)	883
European	Uppsala Longitudinal Study of Adult Men (ULSAM)	Sweden	166 (0)	953 (0)	565
European	Mt Sinai BioMe Biobank Platform (BioMe)	US	132 (26.5)	455 (34.7)	409
European	Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS)	Sweden	111 (41.4)	838 (51.2)	392
European	Estonian Genome Center, University of Tartu (EGCUT-370)	Estonia	80 (48.8)	1768 (51.3)	306
Total Genome-Wide Array			11,645	32,769	28,350
Total Whole Genome Sequence + Genome-Wide Array			12,971	34,100	31,007
Whole Exome Sequencing Studies					
African American	Jackson Heart Study	US	500 (66.6)	526 (63.3)	1,026
African American	Wake Forest School of Medicine Study	US	518 (59.5)	530 (56.0)	1,048
East Asian	Korea Association Research Project	Korea	526 (45.6)	561 (58.5)	1,086
East Asian	Singapore Diabetes Cohort Study; Singapore Prospective Study Program	Singapore (Chinese)	486 (52.1)	592 (61.3)	1,068
European	Ashkenazi	US, Israel	506 (47.0)	355 (56.9)	834
European	Metabolic Syndrome in Men Study (METSIM)	Finland	484 (0)	498 (0)	982
European	Finland-United States Investigation of NIDDM Genetics (FUSION) Study	Finland	472 (42.6)	476 (45.0)	948
European	Kooperative Gesundheitsforschung in der Region Augsburg (KORA)	Germany	97 (44.3)	90 (63.3)	186
European	UK Type 2 Diabetes Genetics Consortium (UKT2D)	UK	322 (45.7)	320 (82.8)	642
European	Malmo-Botnia Study	Finland, Sweden	478 (54.8)	443 (43.8)	920
Hispanic	San Antonio Family Heart Study, San Antonio Family Diabetes/Gallbladder Study, Veterans Administration Genetic Epidemiology Study, and the Investigation of Nephropathy and Diabetes Study Family Component	US	272 (58.8)	218 (58.7)	484
Hispanic	Starr County, Texas	US	749 (59.7)	704 (71.9)	1,452
South Asian	London Life Sciences Population Study (LOLIPOP)	UK (Indian Asian)	531 (14.1)	538 (15.8)	1,068
South Asian	Singapore Indian Eye Study	Singapore (Indian Asian)	563 (44.4)	585 (49.2)	1,148
Total Whole Exome Sequence			6,504	6,436	12,892
Exome Array Studies					
European	ADDITION; Steno Diabetes Centre (SDC); Health06; Health08; Vejle Biobank; Inter99	Denmark	5813 (40.0)	7987 (54.4)	13,458
European	Wellcome Trust Case Control Consortium (UK Type 2 Diabetes Consortium); Young Diabetes Study (YDX); Genetics of Diabetes and Audit Research Tayside Study (GoDARTS); Oxford Biobank; TwinsUK; 1958 Birth Cohort (BC58)	UK	3576 (51.7)	12675 (41.2)	11,156
European	Finland-United States Investigation of NIDDM Genetics (FUSION) Study; Finrisk2007; Metabolic Syndrome in Men Study (METSIM); Dose-Responses to Exercise Training (DR'sEXTRA); D2D2007	Finland	3593 (33.4)	8222 (26.0)	10,001
European	Malmo Diabetes Cohort (MDC); All New Diabetes in Skane (ANDIS)	Sweden	4633(41.0)	5404 (59.5)	9,978
European	Prevalence, Prediction and Prevention of Diabetes (PPP); Diabetes Register in Vaasa (DIREVA)	Finland	2910 (43.7)	4596 (53.7)	7,127
European	Nurses' Health Study (NHS)	US	1413 (100.0)	1695 (100.0)	3,082
European	Health Professionals Follow-up Study (HPFS)	US	1184 (0.0)	1287 (0.0)	2,467
European	The Exeter Family Study of Child Health (EFSOCH)	UK	1446 (39.0)	1567 (52.0)	3,008
European	Kooperative Gesundheitsforschung in der Region Augsburg (KORA)	Germany	933 (45.3)	2705 (51.7)	2,775
European	Estonian Genome Center at the University of Tartu (EGCUT)	Estonia	882 (43.7)	1506 (44.2)	2,225
European	Gene-Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk (GLACIER)	Sweden	960 (47.6)	957 (54.5)	1,917
European	Fenland cohort of the European Prospective Investigation of Cancer (Fen-EPIC)	UK	691(47.0)	1157 (54.5)	1,730
European	The Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS); Uppsala Longitudinal Study of Adult Men (ULSAM)	Sweden	271(16.9)	1791 (23.9)	942
Total Exome Array			28,305	51,549	69,866
Total Whole Exome Sequence + Exome Array			34,809	57,985	82,758

Extended Data Table 2 | Counts and properties of variants identified in sequenced subjects

a

Genomes integrated panel

	SNV	Indel	SV
Variant Type	25.2M (94%)	1.50M (5.6%)	8,876 (0.03%)
N (%total)			
	Coding	Non-coding	
Function	888K (3.3%)	25.8M (97%)	
N (%total)			
	Rare (MAF<0.5%)	Low frequency (0.5<MAF<5%)	Common (MAF>5%)
Frequency spectrum	6.26M (23%)	4.16M (16%)	16.3M (61%)
N (%total)			
	b137	Novel	
dbSNV	14.6M (55%)	12.1M (45%)	
N (%total)			

b

Exome sequence data

	All samples	African-American	East-Asian	European	Hispanic	South-Asian
Samples:	13,008	2,086	2,165	4,579	1,959	2,219
T2D cases	6,504	1,018	1,012	2,359	1,021	1,094
T2D controls	6,436	1,056	1,153	2,182	922	1,123
<i>Excluded from association analysis</i>	<i>68</i>	<i>12</i>	<i>0</i>	<i>38</i>	<i>16</i>	<i>2</i>
Coverage:						
<i>Coding:</i>						
Mean (<i>Mc</i>) per gene	81.7 ±23.7	83.2 ±24.0	84.6 ±23.8	78.6 ±23.3	83.8 ±24.1	78.2 ±23.2
# of genes with <i>Mc</i> <20	368	302	302	351	269	325
<i>Non-coding:</i>						
Mean per gene	59.0 ±21.0	60.9 ±21.5	62.2 ±21.6	57.5 ±20.6	59.2 ±21.2	55.4 ±20.3
# of genes with <i>Mc</i> <20	1,150	738	731	1,102	804	945
Variant annotations:						
Synonymous SNV	627,630	237,430	178,232	192,282	156,231	211,218
Missense SNV	1,110,897	354,797	296,707	327,049	231,351	344,191
Start SNV	2,055	593	523	639	384	583
Nonsense SNV	26,321	7,188	6,668	8,030	4,660	7,339
Frameshift INDEL	26,901	6,605	6,159	7,515	4,155	6,609
Inframe INDEL	11,090	3,471	2,963	3,145	2,068	3,165
3'UTR SNV, INDEL	65,013	24,583	19,149	21,102	16,959	22,177
5'UTR SNV, INDEL	43,965	16,920	13,520	15,562	11,634	15,595
Intron SNV, INDEL	931,449	352,398	270,564	296,970	243,139	314,810
Essential splicing SNV, INDEL	14,286	3,648	3,454	4,108	2,301	3,744
Other splicing SNV, INDEL	128,644	45,876	35,413	38,263	30,301	41,122
Non-coding RNA SNV, INDEL	18,113	7,247	5,996	6,715	5,084	6,706
Intergenic SNV, INDEL	37,345	14,335	11,498	13,614	10,700	12,937
All	3,043,709	1,075,091	850,846	934,994	718,967	990,196
Coding frequency spectrum:						
Rare (MAF<0.5%)	95.79%	83.30%	90.06%	89.19%	84.56%	89.89%
<i>private</i>	77.93%	53.79%	65.47%	51.80%	37.26%	61.55%
<i>cosmopolitan</i>	0.35%	1.80%	3.02%	1.88%	2.24%	1.73%
Low frequency (0.5<MAF<5%)	2.57%	10.36%	4.61%	5.52%	8.21%	5.10%
<i>private</i>	0.17%	1.43%	1.10%	0.26%	0.52%	1.02%
<i>cosmopolitan</i>	0.60%	1.50%	1.54%	1.94%	2.74%	1.62%
Common (MAF>5%)	1.65%	6.35%	5.33%	5.29%	7.23%	5.00%
<i>private</i>	0.09%	0.00%	0.00%	0.00%	0.01%	0.00%
<i>cosmopolitan</i>	1.50%	4.35%	5.17%	4.97%	6.88%	4.86%
Intron/UTR frequency spectrum:						
Rare (MAF<0.5%)	94.09%	78.68%	86.91%	86.17%	81.43%	86.68%
<i>private</i>	74.76%	49.81%	61.36%	45.26%	31.03%	56.96%
<i>cosmopolitan</i>	0.46%	2.07%	3.98%	2.49%	2.66%	2.19%
Low frequency (0.5<MAF<5%)	3.52%	12.57%	5.63%	6.51%	9.43%	6.32%
<i>private</i>	0.25%	1.74%	1.25%	0.29%	0.47%	1.18%
<i>cosmopolitan</i>	0.80%	1.81%	2.11%	2.53%	3.30%	2.17%
Common (MAF>5%)	2.39%	8.76%	7.46%	7.32%	9.14%	7.00%
<i>private</i>	0.15%	0.00%	0.00%	0.01%	0.00%	0.00%
<i>cosmopolitan</i>	2.17%	5.94%	7.26%	6.93%	8.77%	6.81%

a, Variant numbers for the 2,657 individuals with whole-genome sequence data passing QC and included in the association analysis data set. b, Variant numbers are provided for the 13,008 individuals passing initial rounds of QC from which further QC defined the 12,940 subjects included in the association analysis data set. Private refers to variants seen in only a single ancestral group; cosmopolitan to variants seen in all five major ancestry groups.

Extended Data Table 3 | Characterization of variant associations through conditional analysis

Locus	Variant	MAF	Unconditional and conditional association p-values					Interpretation
Non-coding variant associations characterized in 38,738 samples in GoT2D genome wide imputed meta analysis								
IRS1			rs78124264	rs7578326	rs2943640	rs2943641	The association signal rs78124264 and the GWAS SNPs at this locus are distinct. Signals are not extinguished in reciprocal conditional analysis. Previous GWAS signals are not mediated through rs78124264, which represents a distinct association signal at this locus.	
	rs78124264	0.022	8.5x10 ⁻⁵	2.5x10 ⁻⁷	2.5x10 ⁻⁷	2.5x10 ⁻⁷		
	rs7578326 [†]	0.35	1.2x10 ⁻⁷	1.1x10 ⁻⁵	n.d.	n.d.		
	rs2943640 [†]	0.35	2.5x10 ⁻¹¹	n.d.	4.5x10 ⁻¹⁰	n.d.		
	rs2943641 [†]	0.36	9.0x10 ⁻¹²	n.d.	n.d.	1.5x10 ⁻¹⁰		
PPARG			rs79856023	rs1801282	The association signal rs79856023 and the GWAS SNP at this locus are distinct. Signals are not extinguished in reciprocal conditional analysis. Previous GWAS signal is not mediated through rs79856023, which represents a distinct association signal at this locus.			
	rs79856023	0.022	1.2x10 ⁻⁴	9.2x10 ⁻⁷				
	rs1801282 [†]	0.13	1.6x10 ⁻⁶	1.2x10 ⁻⁵				
Coding variant associations characterized in 28,305 cases and 51,549 controls typed on exome array								
PAM			Asp563Gly (PAM)	Ser1207Gly (PPIP5K2)	Association signals for PAM Asp563Gly and PPIP5K2 Ser1207Gly are indistinguishable in reciprocal conditional analysis. Gene biology, as well as previous reports of additional PAM variants associated with T2D in Icelandic cohorts, highlights PAM as the probable transcript at this locus.			
	Asp563Gly	0.054	1.7x10 ⁻⁷	0.24				
	Ser1207Gly	0.054	0.30	1.0x10 ⁻⁶				
MTMR3-ASCC2			Asn960Ser (MTMR3)	Val123Ile (ASCC2)	Asp407His (ASCC2)	Pro423Ser (ASCC2)	Association signals for the MTMR3 and ASCC2 coding variants are indistinguishable in reciprocal conditional analysis. The MTMR3 Asn960Ser variant has the strongest signal, and highlights MTMR3 as the most likely effector transcript at this locus.	
	Asn960Ser	0.083	3.2x10 ⁻⁶	0.022	0.027	0.022		
	Val123Ile	0.083	0.15	2.0x10 ⁻⁵	0.066	0.76		
	Asp407His	0.083	0.18	0.99	1.9x10 ⁻⁵	0.88		
	Pro423Ser	0.083	0.18	0.67	0.98	2.0x10 ⁻⁵		
KCNJ11-ABCC8			Val337Ile (KCNJ11)	Lys23Glu (KCNJ11)	Ala1369Ser (ABCC8)	Association signals for KCNJ11 Val337Ile and Lys23Glu and ABCC8 Ala1369Ser are indistinguishable in reciprocal conditional analysis. The relative causal contributions of the two genes, making up the two components of the sulfonylurea-responsive potassium channel, are indistinguishable on statistical grounds.		
	Val337Ile	0.40	3.4x10 ⁻⁹	0.17	0.049			
	Lys23Glu	0.40	0.48	5.1x10 ⁻⁹	0.082			
	Ala1369Ser	0.40	0.68	0.84	2.3x10 ⁻⁸			
WFS1			Val333Ile	Asn500Asn	Arg611His	rs4689388	Association signals for the WFS1 coding variants are indistinguishable from each other and the previously reported non-coding GWAS SNP at this locus in reciprocal conditional analysis. WFS1 is the likely effector transcript for the non-coding GWAS signal at this locus, although the causal variant in the gene is unclear.	
	Val333Ile	0.30	9.3x10 ⁻¹²	0.024	0.00070	0.0030		
	Asn500Asn	0.41	0.0070	2.0x10 ⁻¹²	0.0049	0.027		
	Arg611His	0.47	0.020	0.62	1.3x10 ⁻¹⁰	0.19		
	rs4689388 [†]	0.43	0.011	0.62	0.024	2.3x10 ⁻¹¹		
CILP2-TM6SF2			Glu167Lys (TM6SF2)	rs10401969	Association signals for TM6SF2 Glu167Lys and the previously reported non-coding GWAS SNP at this locus are indistinguishable from each other in reciprocal conditional analysis. TM6SF2 is the probable effector transcript for the non-coding GWAS signal at this locus, with the effect mediated through Glu167Lys.			
	Glu167Lys	0.082	1.9x10 ⁻⁷	0.52				
	rs10401969 [†]	0.083	0.62	4.2x10 ⁻⁷				
GRB14-COBL1			Asn939Asp (COBL1)	rs13389219	Association signals for COBL1 Asn939Asp and the previously reported non-coding GWAS SNP at this locus are partially correlated. The association signal for the GWAS signal is not entirely extinguished in reciprocal conditional analysis. COBL1 is a candidate effector transcript for the GWAS signal at this locus.			
	Asn939Asp	0.12	4.7x10 ⁻¹¹	3.00x10 ⁻⁵				
	rs13389219 [†]	0.39	7.0x10 ⁻⁵	1.9x10 ⁻¹⁰				
Coding variant associations characterized in 44,414 samples in GoT2D genome wide imputed meta analysis								
THADA			Cys1605Tyr	rs10203174	Association signals THADA Cys1605Tyr and the GWAS SNP are partially correlated. The association signal for the GWAS SNP is not entirely extinguished in reciprocal conditional analysis. THADA is a candidate effector transcript for the GWAS signal at this locus.			
	Cys1605Tyr	0.10	0.00035	0.92				
	rs10203174 [†]	0.10	0.0063	5.7x10 ⁻⁶				
RREB1			Asp1171Asn	rs9502570	The association signals of RREB1 Asp1171Asn and the GWAS SNP at this locus are distinct. The association signal is not extinguished in reciprocal conditional analysis. Previous GWAS signal is not mediated through RREB1 Asp1171Asn. RREB1 Asp1171Asn represents a distinct association signal at this locus.			
	Asp1171Asn	0.11	0.0018	0.0017				
	rs9502570 [†]	0.28	0.0037	0.0042				

For each locus, significantly associated SNVs are presented. Unconditional *P* values are given in italics, and conditional *P* values are shown for each pair of SNVs (*P* values are for SNVs in the Variant column, with SNVs listed in header included as covariates in association analysis). The *IRS1* and *PPARG* non-coding associations were characterized using exact conditional analysis in 38,738 samples from the GoT2D genome-wide imputed meta-analysis. Conditional analysis for coding variant associations was, for most loci, restricted to the exome array genotypes (28,305 cases, 51,549 controls). At *THADA* and *RREB1*, neither the non-coding lead GWAS SNVs nor close proxies were typed on the exome array, so approximate conditional analyses were undertaken using GCTA in 44,414 samples from the GoT2D genome-wide imputed meta-analysis (see Methods). For several of these loci, unconditional association *P* values for these loci do not reach genome-wide significance as sample sizes are smaller. At the *GPSM1* locus, the previously reported GWAS SNV was not available on exome array and too poorly imputed in the GoT2D meta-analysis to allow meaningful inference.

*Conditional analysis was performed once for rs78124264 with all three previously known GWAS variants included as covariates.

[†]Non-coding GWAS lead variant. n.d., not determined.

Extended Data Table 4 | Testing for synthetic associations across GWAS-identified T2D loci

		Index SNV association signal			Synthetic association by missense variants				Synthetic association by all low-frequency and rare variants across 5Mb region					
		Index SNV association signal before inclusion of missense variants			Index SNV association signal after inclusion of missense variants				Index SNV association after inclusion of single best variant				Testing groups of low-frequency and rare variants	
Gene	Index SNV	MAF	OR [95% interval]	p-value	Number Missense Variants	OR [95% interval]	p-value	Relative likelihood of LF model	Best LF Variant	MAF	OR [95% interval]	p-value	n ₁	n ₂
TCF7L2	10:114758349	0.27	1.75 [1.54-1.99]	2.80×10 ⁻¹⁸	6	1.73 [1.52-1.97]	2.33×10 ⁻¹⁷	1.8×10 ⁻¹⁷	10:114787948	1.6%	1.72 [1.51-1.95]	1.62×10 ⁻¹⁶	>50	35
ADCY5	3:123065778	0.19	0.69 [0.60-0.79]	1.12×10 ⁻⁷	13	0.70 [0.61-0.81]	9.00×10 ⁻⁷	9.7×10 ⁻⁸	3:123096056	2.5%	0.71 [0.61-0.82]	3.04×10 ⁻⁶	13	6
IRS1	2:227093745	0.36	0.76 [0.68-0.86]	2.80×10 ⁻⁶	5	0.77 [0.69-0.86]	4.30×10 ⁻⁶	4.5×10 ⁻⁷	2:226993370	1.7%	0.78 [0.70-0.88]	2.19×10 ⁻⁵	12	6
KCNQ1	11:2847069	0.45	0.78 [0.70-0.87]	1.22×10 ⁻⁵	>50	0.84 [0.75-0.94]	2.07×10 ⁻³	1.0×10 ⁻⁷	11:2825279	4.7%	0.81 [0.71-0.91]	3.19×10 ⁻⁴	16	6
CDC123-CAMK1D	10:12307894	0.25	1.33 [1.17-1.52]	1.19×10 ⁻⁵	4	1.30 [1.13-1.50]	2.06×10 ⁻⁴	7.1×10 ⁻⁵	10:12325477	3.8%	1.29 [1.12-1.48]	3.03×10 ⁻⁴	10	5
CDKN2A-CDKN2B	9:22137685	0.28	1.28 [1.14-1.45]	4.52×10 ⁻⁵	4	1.27 [1.13-1.43]	9.28×10 ⁻⁵	4.3×10 ⁻⁵	9:22133773	3.5%	1.25 [1.10-1.41]	5.98×10 ⁻⁴	22	7
IGF2BP2	3:185511687	0.32	1.25 [1.11-1.41]	1.65×10 ⁻⁴	14	1.21 [1.07-1.36]	2.12×10 ⁻³	3.0×10 ⁻⁴	3:185550500	4.1%	1.20 [1.07-1.36]	2.91×10 ⁻³	8	3
KLHDC5	12:27965150	0.17	0.76 [0.66-0.88]	2.19×10 ⁻⁴	3	0.77 [0.66-0.89]	4.45×10 ⁻⁴	1.2×10 ⁻³	12:27832062	2.0%	0.80 [0.68-0.92]	3.04×10 ⁻³	10	4
SLC30A8	8:118184783	0.33	0.81 [0.72-0.91]	2.95×10 ⁻⁴	2	0.81 [0.72-0.91]	3.73×10 ⁻⁴	0.02	8:117964024	2.2%	0.83 [0.73-0.93]	1.23×10 ⁻³	17	6
CDKAL1	6:20694884	0.18	1.28 [1.11-1.48]	6.05×10 ⁻⁴	1	1.28 [1.11-1.48]	7.57×10 ⁻⁴	0.007	6:20718780	2.8%	1.23 [1.06-1.43]	7.71×10 ⁻³	9	3

Gene names refer to protein-coding transcript(s) closest to the index SNV. Reported index SNVs are the previously reported GWAS variants (in European populations) with the strongest association signal in the GoT2D sequencing data ($n = 2,657$). Relative likelihoods are based on causal models with only the chosen low-frequency and rare missense variants, relative to models with only the GWAS index SNV, assessed using the Akaike Information content (AIC) of each regression model, calculated as $\exp[(\text{AIC}_{\text{index}} - \text{AIC}_{\text{low-frequency or rare}})/2]$. n_1 , number of low-frequency or rare variants required for the residual odds ratio at the GWAS index SNV, after joint conditioning on the low-frequency and rare variants, to switch direction of effect. n_2 , number of low-frequency or rare variants required for the association P value remaining at the GWAS index SNV, after joint conditioning on the low-frequency and rare variants, to exceed 0.05.

Structural basis of potent Zika–dengue virus antibody cross–neutralization

Giovanna Barba-Spaeth^{1,2*}, Wanwisa Dejnirattisai^{3*}, Alexander Rouvinski^{1,2*}, Marie-Christine Vaney^{1,2*}, Iris Medits⁴, Arvind Sharma^{1,2}, Etienne Simon-Lorière^{5,6}, Anavaj Sakuntabhai^{5,6}, Van-Mai Cao-Lormeau⁷, Ahmed Haouz^{8,9}, Patrick England^{9,10}, Karin Stiasny⁴, Juthathip Mongkolsapaya^{3,11}, Franz X. Heinz⁴, Gavin R. Screaton³ & Félix A. Rey^{1,2}

Zika virus is a member of the Flavivirus genus that had not been associated with severe disease in humans until the recent outbreaks, when it was linked to microcephaly in newborns in Brazil and to Guillain–Barré syndrome in adults in French Polynesia. Zika virus is related to dengue virus, and here we report that a subset of antibodies targeting a conformational epitope isolated from patients with dengue virus also potently neutralize Zika virus. The crystal structure of two of these antibodies in complex with the envelope protein of Zika virus reveals the details of a conserved epitope, which is also the site of interaction of the envelope protein dimer with the precursor membrane (prM) protein during virus maturation. Comparison of the Zika and dengue virus immunocomplexes provides a lead for rational, epitope–focused design of a universal vaccine capable of eliciting potent cross–neutralizing antibodies to protect simultaneously against both Zika and dengue virus infections.

Zika virus (ZIKV) is an arthropod-borne enveloped virus belonging to the Flavivirus genus in the family *Flaviviridae*, which also includes the human pathogenic yellow fever, dengue, West Nile and tick-borne encephalitis viruses¹. Flaviviruses have two structural glycoproteins, prM and E (for precursor membrane and envelope proteins, respectively), which form a heterodimer in the endoplasmic reticulum (ER) of the infected cell and drive the budding of spiky immature virions into the ER lumen. These particles transit through the cellular secretory pathway, during which the *trans*-Golgi-resident protease furin cleaves prM. This processing is required for infectivity, and results in the loss of a large fragment of prM and reorganization of E on the virion surface. The mature particles have a smooth aspect, with 90 E dimers organized with icosahedral symmetry in a ‘herringbone’ pattern^{2,3}.

Three-dimensional cryo-electron microscopy (cryo-EM) structures of the mature ZIKV particles have recently been reported to near atomic resolution (3.8 Å)^{4,5}, showing that the virus has essentially the same organization as the other flaviviruses of known structure, such as dengue virus (DENV)³ and West Nile virus^{6,7}. The E protein is about 500 amino acids long, with the 400 N-terminal residues forming the ectodomain essentially folded as β -sheets with three domains, named I, II and III, aligned in a row with domain I at the centre. The conserved fusion loop is at the distal end of the rod in domain II, buried at the E dimer interface. At the C terminus, the E ectodomain is followed by the ‘stem’, featuring two α -helices lying flat on the viral membrane (the stem helices), which link to two C-terminal transmembrane α -helices. The main distinguishing feature of the ZIKV virion is an insertion within a glycosylated loop of E (the ‘150’ loop), which protrudes from the mature virion surface^{4,5}.

Flaviviruses have been grouped into serocomplexes based on cross-neutralization studies with polyclonal immune sera⁸. The E protein is the main target of neutralizing antibodies, and is also the

viral fusogen; cleavage of prM allows E to respond to the endosomal pH by undergoing a large-scale conformational change that catalyses membrane fusion and releases the viral genome into the cytosol. Loss of the precursor fragment of prM lets the E protein fluctuate from its tight packing at the surface of the virion, transiently exposing otherwise buried surfaces. One surface exposed by this ‘breathing’ is the fusion-loop epitope (FLE), which is a dominant cross-reactive antigenic site⁹. Although antibodies to this site can protect by complement-mediated mechanisms, as shown in a mouse model for West Nile virus¹⁰, they are poorly neutralizing and lead to antibody-dependent enhancement^{11–15}, thereby aggravating Flavivirus pathogenesis and complicating the development of safe and effective vaccines.

We recently reported the functional and structural characterization of a panel of antibodies isolated from patients with dengue disease^{13,16}. Most of these antibodies target the FLE, but others target a quaternary site readily accessible at the exposed surface of the E protein on the virion, at the interface between the two E subunits in the dimer. These broadly neutralizing antibodies (bnAbs), termed EDE for E-dimer epitope, potently neutralize all four DENV serotypes. Their binding site is conserved across serotypes because it is also the interaction site of prM with E dimers during transport of the immature virus particles through the Golgi apparatus of the cell. There were two subsets of EDE antibodies, characterized by a differential requirement for glycosylation on the 150 loop for binding. The EDE1 bnAbs bind better in the absence of glycan, whereas EDE2 bnAbs bind better when the glycan is present.

In this study, we show that the EDE bnAbs neutralize ZIKV as potently as they neutralize DENV. We also find that the FLE antibodies, which neutralize DENV although not as potently as the EDE bnAbs, do not neutralize ZIKV at concentrations up to 1 μ M in spite of a high affinity for the recombinant ZIKV E protein. We further describe the

¹Institut Pasteur, Unité de Virologie Structurale, Département de Virologie, F-75724 Paris Cedex 15, France. ²CNRS UMR 3569 Virologie, F-75724 Paris Cedex 15, France. ³Division of Immunology and Inflammation, Department of Medicine, Hammersmith campus, Imperial College London W12 0NN, UK. ⁴Department of Virology, Medical University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria. ⁵Institut Pasteur, Unité de Génétique fonctionnelle des maladies infectieuses, Département de Génomes et Génétique, F-75724 Paris Cedex 15, France. ⁶CNRS URA 3012, F-75724 Paris Cedex 15, France. ⁷Unit of Emerging Infectious Diseases, Institut Louis Malardé, 98713 Papeete, Tahiti, French Polynesia. ⁸Institut Pasteur, Plateforme de Cristallographie, CiTech, Département de Biologie Structurale et Chimie, F-75724 Paris Cedex 15, France. ⁹CNRS UMR 3528, F-75724 Paris Cedex 15, France. ¹⁰Institut Pasteur, Plateforme de Biophysique des Macromolécules et de leurs Interactions, CiTech, Département de Biologie Structurale et Chimie, F-75724 Paris Cedex 15, France. ¹¹Dengue Hemorrhagic Fever Research Unit, Office for Research and Development, Siriraj Hospital, Faculty of Medicine, Mahidol University, Bangkok, 10700 Thailand.

*These authors contributed equally to this work.

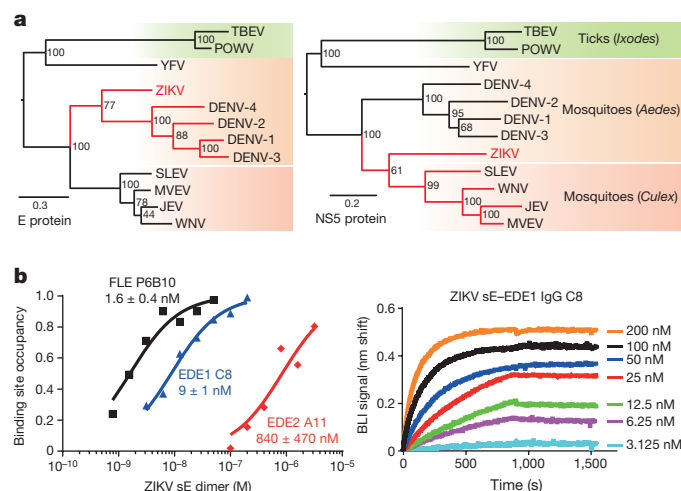


Figure 1 | ZIKV and DENV E protein phylogeny and reactivity with DENV-elicited antibodies. **a**, Phylogenetic trees of the main human pathogenic flaviviruses based on the amino acid sequences of the E protein (left) and of the polymerase NS5 protein (right). The arthropod vectors are differentiated by the background colour. JEV, Japanese encephalitis virus; MVEV, Murray Valley encephalitis virus; POWV, Powassan virus; SLEV, Saint Louis encephalitis virus; TBEV, tick-borne encephalitis virus; YFV, yellow fever virus; WNV, West Nile virus. **b**, ZIKV sE reactivity with human recombinant full-length antibodies FLE P6B10, EDE1 C8 and EDE2 A11. Left, binding properties were monitored by biolayer interferometry (BLI) on Octet-Red (ForteBio). The normalized response values expressed as fraction of binding site occupancy are plotted against concentrations of ZIKV sE dimer shown at logarithmic scale. Lines denote global curve fits used for K_d evaluation (see Extended Data Fig. 1a for linear concentration range showing concentration-dependent saturation fits). Normalized response values were deduced from individual sensorgrams showing the binding properties for EDE1 C8 measured at different ZIKV sE concentrations.

crystal structures of the ZIKV E protein dimer alone and in complex with EDE1 C8 and EDE2 A11, identifying their binding determinants.

A ZIKV–DENV super serogroup

Phylogenetic analyses of the main human pathogenic flaviviruses using the amino acid sequences of the viral RNA polymerase NS5 indicate a clustering of ZIKV with the group of mosquito-borne encephalitic viruses. The clustering is different when the amino acid sequences of the E protein are considered, with ZIKV branching with the DENV group (Fig. 1a). If the sequence clustering extends to the antigenic surface of E, antibodies that cross-react with several DENV serotypes should also bind ZIKV E. To test this hypothesis, we used biolayer interferometry to study the binding properties of a poorly neutralizing, cross-reactive FLE antibody and the potentially neutralizing EDE

antibodies for recombinant, soluble ZIKV E ectodomain (ZIKV sE) produced in insect cells (see Methods). The FLE antibody (P6B10) bound nearly 10-fold more tightly than did EDE1 C8 (apparent dissociation constant (K_d) values of 1.5 nM versus 9 nM), and nearly 1,000 times more tightly than EDE2 A11 (Fig. 1b and Extended Data Fig. 1a). Consistent with their affinities, we could isolate a complex of ZIKV sE with a C8 Fab by size-exclusion chromatography, but not with an A11 Fab (Extended Data Fig. 1b).

Neutralization assays in African green monkey kidney (Vero) cells using these and other members of the three antibody subsets, showed that the EDE1 antibodies strongly neutralize ZIKV, whereas the EDE2 antibodies were at least one log order less potent. Despite its strong affinity, P6B10 did not neutralize ZIKV in the concentration range used, nor did either of the two other FLE antibodies tested (Fig. 2). The EDE1 bnAbs neutralized best the ZIKV African strain HD78788, with a half-maximum inhibitory concentration (IC_{50}) of about 0.1 nM. This strain has over the years been cell-culture-adapted and passaged in suckling mice brain and lacks E glycosylation. The IC_{50} against the PF13 strain, isolated in French Polynesia in 2013 and in which the E protein is glycosylated at position 154, was in the nanomolar range and comparable to, or lower than, that against the four DENV serotypes (Table 1). The EDE2 bnAbs showed no difference in neutralization of the two strains, suggesting that the presence of the N154 glycan in the ZIKV E protein did not enhance the interaction.

The ZIKV–EDE bnAbs immune complexes

We crystallized unliganded ZIKV sE and complexes of ZIKV sE with EDE1 C8 and EDE2 A11 with single-chain Fv (scFv) and Fab fragments, respectively (Extended Data Table 1). In the structure of unliganded ZIKV sE, the 150 loop is ordered, unlike the unglycosylated 150 loop in the recently determined structure of the protein produced in bacteria and refolded *in vitro*¹⁷. In contrast to our insect-cell-secreted protein, which is a dimer (Extended Data Fig. 1b), the refolded protein was reported to be monomeric in solution, suggesting that the glycan may help to structure the loop and promote sE dimerization.

The antibodies recognize a quaternary epitope in the ZIKV sE dimer in the same way that they recognize the DENV serotype 2 (DENV-2) sE dimer described earlier¹⁶. The amino acid residues participating in the contacts, for both the ZIKV and DENV-2 structures, are shown in Extended Data Fig. 2. As expected, the pattern is very similar, with the few differences highlighted in red frames in Extended Data Fig. 2b. Both epitopes in the sE dimer are occupied in the case of the complex with C8 (Fig. 3a), whereas only one is occupied in the case of A11 (Fig. 4a). Inspection of the crystal environment showed that a second Fab could not be docked at this position without clashing with neighbouring complexes in the crystal. This observation indicates that crystal growth selected for incorporation of sE dimers with a single Fab bound, which is facilitated by the low affinity of A11.

The bnAbs dock on ZIKV sE at different angles than they do on DENV-2 sE (see insets in Figs 3a and 4a). In the case of the C8 complex,

Table 1 | 50% FRNT values of EDE1, EDE2 and FLE antibodies tested against ZIKV and DENV-1–4

		ZIKV		DENV			
		50% FRNT (nM)		50% FRNT (nM)			
	Epitope	PF13	HD78788	DENV-1	DENV-2	DENV-3	DENV-4
752-2-C8	EDE1	0.095 (±0.026)	0.015 (±0.003)	0.39 (±0.21)	0.24 (±0.06)	0.64 (±0.08)	1.13 (±0.14)
753(3)-C10	EDE1	0.063 (±0.016)	0.013 (±0.025)	0.54 (±0.04)	0.18 (±0.02)	1.89 (±0.79)	0.08 (±0.03)
752-2-B2	EDE1	1.062 (±0.362)	0.021 (±0.004)	0.32 (±0.05)	0.23 (±0.02)	0.22 (±0.09)	0.44 (±0.14)
747(4)-A11	EDE2	0.904 (±0.191)	0.506 (±0.102)	0.11 (±0.01)	0.07 (±0.03)	0.11 (±0.02)	7.79 (±3.19)
747(4)-B7	EDE2	4.31 (±1.47)	1.17 (±0.180)	0.10 (±0.01)	0.11 (±0.02)	0.12 (±0.03)	93.19 (±19.15)
747-C4	EDE2	102 (±25.6)	11.6 (±2.6)	0.23 (±0.02)	0.08 (±0.01)	0.11 (±0.01)	0.12 (±0.05)
758-P6B10	FLE	No neut.	No neut.	1.85 (±0.44)	4.97 (±0.28)	9.40 (±2.83)	7.47 (±1.65)
749-B12	FLE	No neut.	No neut.	0.43 (±0.12)	0.73 (±0.20)	1.04 (±0.31)	1.80 (±0.64)
750-2-C5	FLE	No neut.	No neut.	1.08 (±0.21)	0.76 (±0.46)	1.40 (±0.25)	2.30 (±0.02)

FRNT, focus reduction neutralization test. Data shown are mean values with s.e.m. in parentheses.

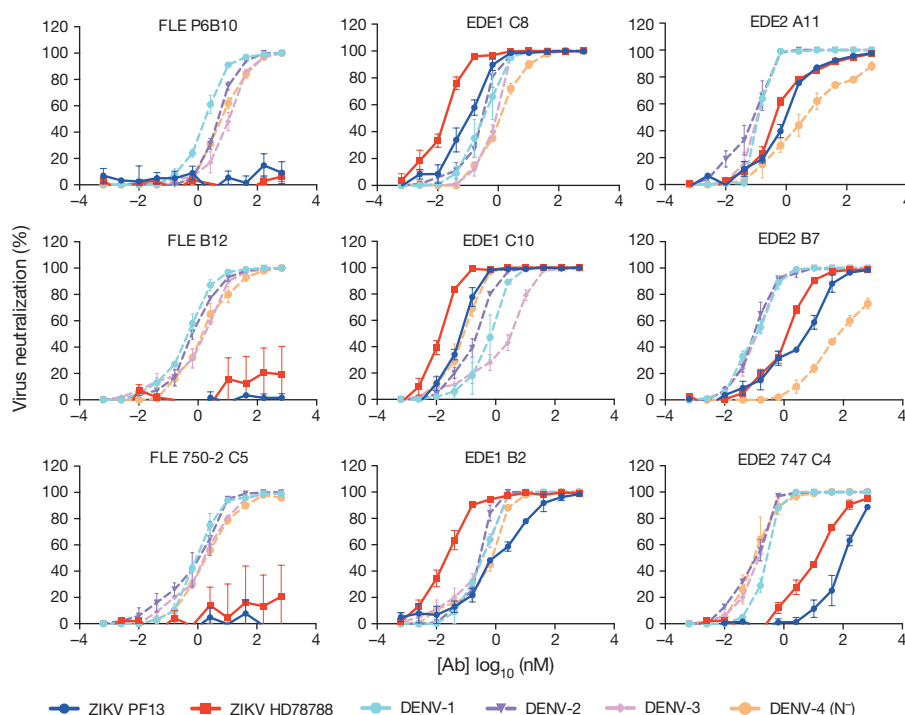


Figure 2 | Neutralization curves using three antibodies each from the three subsets FLE, EDE1 and EDE2. Results represent the mean \pm s.e.m. of four independent experiments performed in triplicate for PF13 and in duplicate for HD78788 strains. The two ZIKV strains are in bright colours, blue and red, respectively. The neutralization data for the four DENV

serotypes (dotted lines in pale colours) were taken from ref. 13, and are given here for comparison. The corresponding IC_{50} values are provided in Table 1. Note that the DENV-4 strain used was a natural isolate lacking the N153 glycosylation site (N^-).

the difference in docking results mainly from an altered curvature of the sE dimer. We note that the conformation of ZIKV sE in complex with the antibodies is very similar to the one it adopts on the virus particle, with roughly 1.5 Å root mean square deviation (r.m.s.d.) for 790 C α atoms (see Extended Data Table 2). The unbound ZIKV sE crystallized here displays a more distant conformation (2.5 Å r.m.s.d. when comparing to both virion ZIKV E and either of the sE antibody complexes), suggesting that the antibodies stabilize a conformation close to that on the viral particle. By contrast, the same comparisons done for DENV-2 sE, alone or in complex with the bnAbs, result in r.m.s.d. values of 5–7 Å with respect to the E conformation on the DENV virion observed by cryo-EM³. For comparison, superposition of the ectodomain of virion E from ZIKV⁵ and DENV-2 (ref. 3) results in a similar 1.5 Å r.m.s.d. value, indicating that they are presented roughly in the same way, but that DENV sE is more deformable in solution. This malleability may reflect the conformational breathing reported for DENV E¹⁸. Instead, ZIKV sE remains in a similar conformation in the absence of the interactions with the underlying stem α -helices and with the M protein (the membrane-anchored remnant of prM after furin cleavage) on the virion, in line with the higher stability of the ZIKV particles described recently⁴.

EDE1 C8 complex

The total buried surface area of EDE1 C8 in the complex with ZIKV sE is about 900 Å², compared to about 1,300 Å² in the DENV-2 sE complex (Extended Data Table 3). Figure 3d shows the conservation of the epitope, and Fig. 3e and f compare the C8 footprint on ZIKV and DENV-2 sE. The DENV-specific glycan at position N67, which is ordered in the DENV-2 sE structure (Fig. 3c), accounts for around two-thirds of the overall difference in footprint area. The N67 glycan interacts with the framework region 2 of the heavy chain (FRH2), and its absence in ZIKV sE shows that these contacts are not essential for binding. The key cluster of interactions that is maintained is centred on β -strand *b* of domain II, with side chains from

complementarity determining regions (CDRs) H2, H3 and L3 recognizing all the available hydrogen bond donors (main-chain carbonyls) and acceptors (main-chain carbonyls) of the *bdc* β -sheet edge (Fig. 3b, c). In addition, the fusion loop main chain (which contains several glycine residues) and the disulfide bond between Cys74 and Cys105 are framed by aromatic side chains of the CDRs L1 and L3 (see also Extended Data Fig. 3). Residues from these two CDRs also recognize strictly conserved side chains of the fusion loop (Arg99) or nearby residues (Gln77).

Across the dimer interface, and as in the complex with DENV-2, the 150 loop is partially disordered, with no detectable density for the N154 glycan (Fig. 3a and Extended Data Fig. 3d). As shown in Extended Data Fig. 3, the interacting residues across the dimer interface are different, reflecting the more limited sequence conservation in these regions of the E protein: in the DENV-2 sE complex, these contacts are with residues from β -strands A and B of domain III, but in ZIKV they mainly involve Lys373 from β -strand E interacting with CDRs L1 and L2, via a network of direct or water-mediated hydrogen bonds (Extended Data Fig. 3b, c). Similarly, several charged residues in domain I and from the nearby *kl* loop of domain II across the interface, contribute to the binding and interact with the heavy chain CDRs H2 and H3 (Extended Data Fig. 3e, f). All of the polar interactions between C8 and ZIKV sE are listed in Extended Data Tables 4 and 5, and the electrostatic surface of the epitope is shown in Extended Data Fig. 4, left panel. In summary, these observations identify the conserved cluster of contacts with the *b* strand and the fusion loop in domain II as the main binding determinants of C8, with additional contacts from across the dimer interface—or from the N67 glycan in DENV—further stabilizing but not determining the interaction.

EDE2 A11 complex

Extended Data Fig. 4 compares the footprint of C8 and A11 on ZIKV sE, together with the surface electrostatic potential of the complexes, which shows a strong basic patch on sE in the C8 complex due to

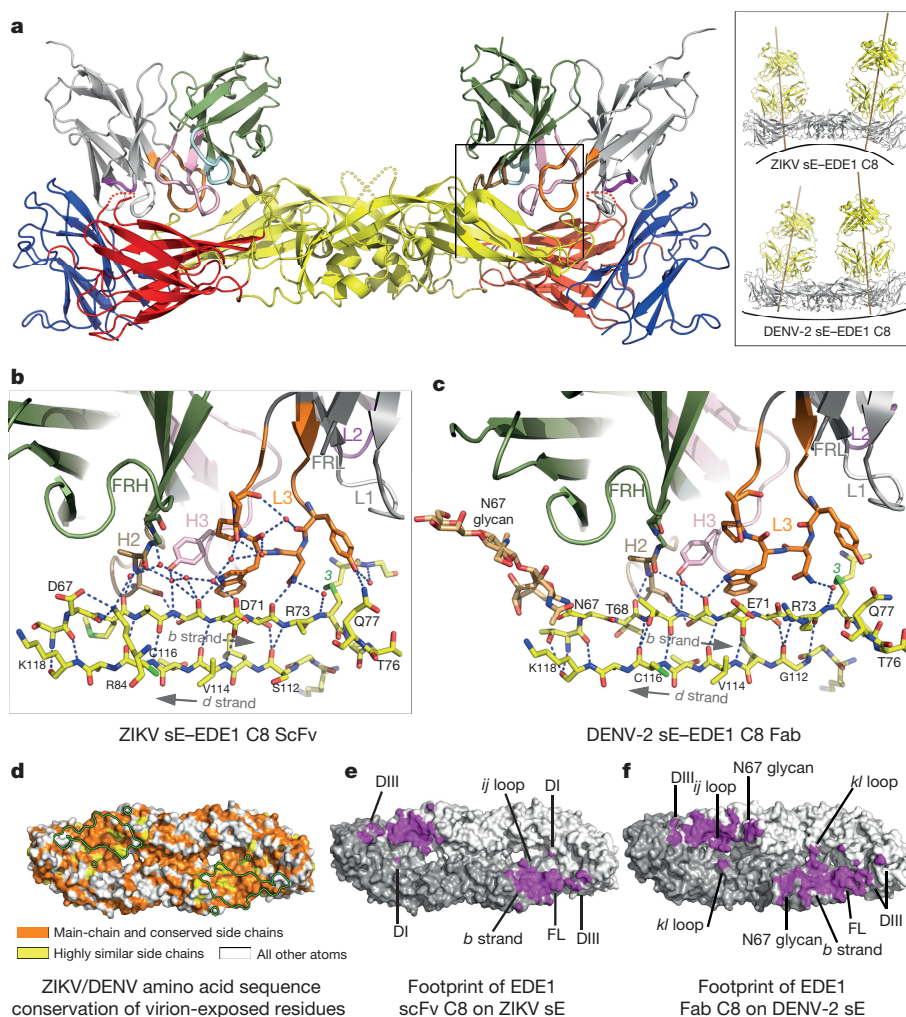


Figure 3 | EDE1 C8-ZIKV sE complex. **a**, Overall view of the complex, with the sE moiety coloured by domains (I, II and III in red, yellow and blue, respectively); the antibodies in grey and dark green for light and heavy chains, respectively. The CDRs are coloured (H1, light blue; H2, sand; H3, pink; L1, light grey; L2, magenta; L3, orange). The inset shows a comparison with the corresponding DENV-2 complex. For clarity, the variable region of the C8 Fab fragment of the DENV-2 sE-C8 complex was superposed on the C8 scFv in complex with ZIKV sE to draw the Fab axes and show the docking angles better. **b**, Zoom of the ZIKV sE-C8 interaction to show the recognition of the *b* strand. Hydrogen bonds are shown as dotted lines and immobilized water molecules as red spheres. **c**, Same region on the DENV-2 sE-C8 Fab complex. Note that the N67

glycan on DENV also interacts with the antibody. **d**, The footprint of EDE1 C8 is outlined on the ZIKV sE dimer shown in surface representation (looking from outside the virion) coloured according to conservation of surface-exposed amino acids. Atoms from the main-chain and conserved side chains are orange, highly similar side chains are yellow, and all the others are white. **e**, **f**, Footprints of EDE1 C8 on a surface representation of ZIKV sE (**e**) and DENV-2 sE (**f**) shown in purple. FL, fusion loop. The two protomers of sE in the dimer are in light and dark grey. Relevant antigenic sE regions are labelled. Note the more confined interacting surface in the ZIKV sE dimer than in DENV-2, for example, N67 glycan is absent in ZIKV sE.

the disorder of the 150 loop. As shown in Extended Data Fig. 5, C8 would clash with the glycan had the loop remained in place, as was the case in the complex with DENV-2 sE¹⁶. In the A11 complex, the 150 loop remains in the same conformation as in the cryo-EM structures of the virion (Extended Data Fig. 5a) and in the X-ray structure of glycosylated unliganded sE reported here. In the DENV-2 sE-A11 complex, the glycan is recognized by an α -helix in the long CDR H3 loop (Fig. 4e). The difference in length in the 150 loop of E in ZIKV compared to DENV shifts the glycan position by about 6–7 Å, such that it cannot make the same interactions with the CDR H3 α -helix (Fig. 4d, e and Extended Data Fig. 5b). As a consequence, the A11 antibody docks at a different angle on ZIKV sE than it does on DENV-2 sE, even accounting for the difference in sE dimer curvature (Fig. 4a, inset). The contacts along the *b* strand are preserved (Fig. 4b, c). Compared to C8, the *b* strand is recognized only along half its length (residues 71 and 73), whereas C8 recognizes it all along, from residue 68 (or from 67 in DENV).

Discussion

Our results identify the structural details of a quaternary epitope that provides a previously unrecognized link of potent cross-neutralization between Zika and dengue viruses, and thus identifies an antigenic Flavivirus cluster beyond the traditional serocomplexes. This relationship defines a super serogroup on the basis of strong cross-neutralization through a conserved epitope that had not been recognized using polyclonal sera⁸. This finding thus introduces the possibility of developing a universal vaccine protecting against all the viruses from this group.

Vaccine design against dengue virus has been hampered by the heterogeneity of DENV particles and the need to use polyvalent formulas to immunize against all four serotypes^{19,20}. One feature of DENV is that it undergoes incomplete furin maturation cleavage of prM in many cell types, giving rise to heterogeneous mosaic particles with an immature-like spiky patch on one side and a smooth mature-like region on the opposite side²¹. These particles are infectious, as they can fuse with

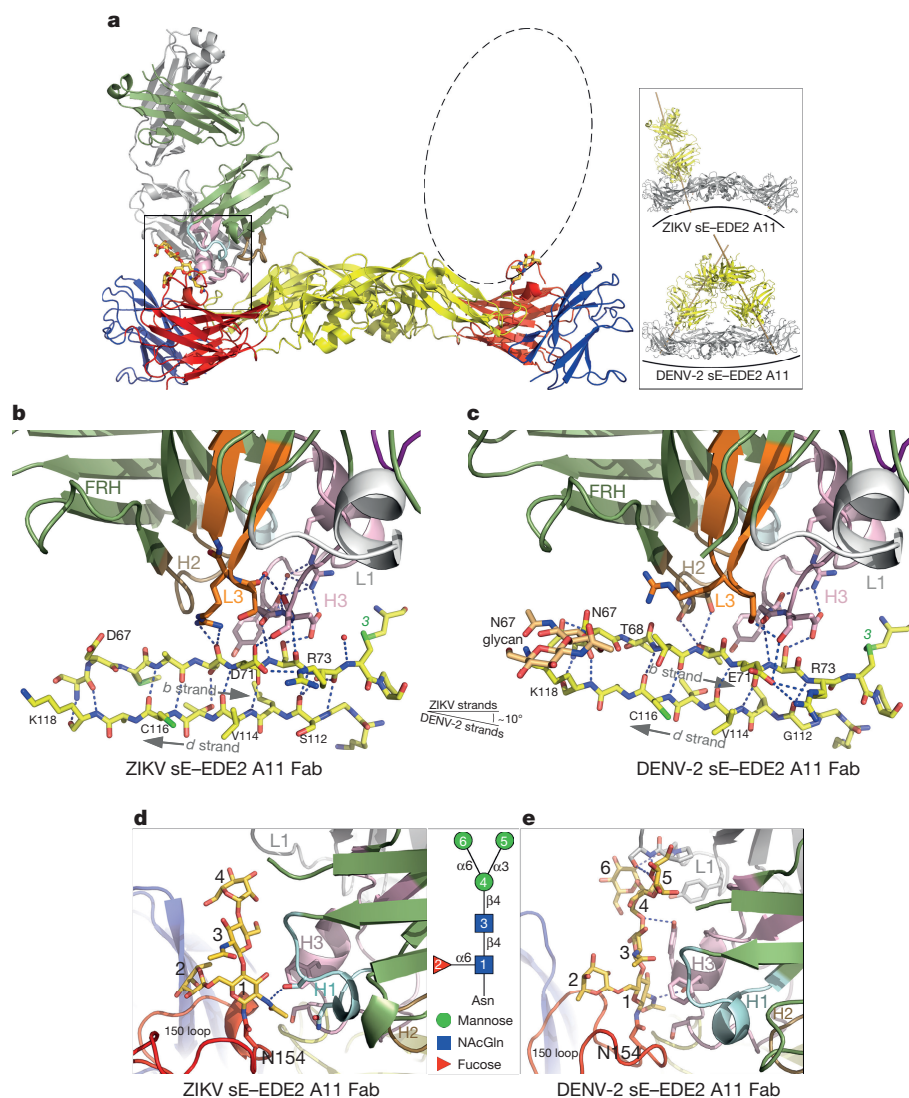


Figure 4 | EDE2 A11–ZIKV sE complex. Colour coding is as in Fig. 3. **a**, Overall view of the complex, with only one Fab bound per sE dimer, owing to crystal packing. The dashed ellipse represents the position of the missing A11 Fab. The inset compares the angle of binding to the sE dimer in ZIKV and in DENV-2. **b**, **c**, Interactions at the *b* strand in ZIKV (left) and in DENV-2 (right). Note the different angle of the *b* strand with

the cellular membrane through the smooth, mature side. Because the FLE is exposed in immature regions²², most of the antibody response in DENV-infected patients is directed against it²³. These cross-reactive antibodies coat the particles on the immature side²² but neutralize only weakly, because they can bind the mature side only when the E protein ‘breathes’^{24–26}. A recently published structure of monomeric ZIKV sE in complex with an FLE-specific monoclonal mouse antibody of low neutralizing activity indeed shows that its binding site would be occluded in the dimeric E protein on mature infectious virions¹⁷. The observation that P6B10 and other FLE antibodies still neutralize DENV¹³ suggests that E in the mature patches on DENV spends more time in conformations that expose the FLE than does E in those patches on ZIKV. This inference is consistent with the higher thermal stability of ZIKV reported recently⁴.

Our results suggest that the epitope targeted by the EDE1 bnAbs is better suited for developing an epitope-focused vaccine for viruses in the ZIKV/DENV super serogroup than is the FLE, which induces poorly neutralizing and strongly infection-enhancing antibodies^{12–14}. The EDE1 is also better suited than the related EDE2 epitope: although the EDE1 bnAbs require an E dimer to bind, the actual binding

respect to the antibody (the antibody is in exactly the same orientation in both panels). **d**, **e**, Zoom of the glycan on the 150 loop for ZIKV sE (**d**) and for DENV-2 sE (**e**), with sugar residue numbers described in the key. The CDR H3 helix is too far to make interactions with the glycan, as is the case in the DENV-2 structure (see Extended Data Figs 2b and 5b).

determinants are centred on the *b* strand and on the highly conserved, E-dimer-exposed elements of the fusion loop, as shown by the comparison between their binding to DENV-2 and ZIKV sE. The fact that EDE2 bnAbs rely heavily on their contact points on the adjacent subunit—on the variable 150 loop in which glycosylation is not always present—is a drawback, as demonstrated by their poor affinity (Fig. 1) and their strong induction of antibody-dependent enhancement¹².

Targeting the *b* strand and the E-dimer-exposed elements of the fusion loop appears as a powerful alternative to the multi-immunogen approaches against the DENV cluster that have had limited success in clinical trials²⁷. As the E protein polypeptide chain displays neither insertions nor deletions in the region of the *b* strand for any medically relevant Flavivirus, this region presents a low risk of inducing escape mutations, most likely because it is also the interacting site with prM during virus maturation. Finally, in a more immediate application, our study also suggests that the EDE1 antibodies, perhaps carrying the ‘LALA’ mutation²⁸ if effector functions are to be avoided, could be useful for immune prophylaxis for pregnant women at risk of contracting ZIKV infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 May; accepted 17 June 2016.

Published online 23 June 2016.

- Lindenbach, B., Murray, C., Thiel, H. & Rice, C. *Flaviviridae: The Viruses and Their Replication* 6th edn, Vol. 1 1101–1152 (Lippincott Williams & Wilkins, 2013).
- Kuhn, R. J. *et al.* Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**, 717–725 (2002).
- Zhang, X. *et al.* Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nat. Struct. Mol. Biol.* **20**, 105–110 (2013).
- Kostyuchenko, V. A. *et al.* Structure of the thermally stable Zika virus. *Nature* **533**, 425–428 (2016).
- Sirohi, D. *et al.* The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **352**, 467–470 (2016).
- Mukhopadhyay, S., Kim, B. S., Chipman, P. R., Rossmann, M. G. & Kuhn, R. J. Structure of West Nile virus. *Science* **302**, 248 (2003).
- Zhang, W., Kaufmann, B., Chipman, P. R., Kuhn, R. J. & Rossmann, M. G. Membrane curvature in flaviviruses. *J. Struct. Biol.* **183**, 86–94 (2013).
- Calisher, C. H. *et al.* Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. *J. Gen. Virol.* **70**, 37–43 (1989).
- Stiasny, K., Kiermayr, S., Holzmann, H. & Heinz, F. X. Cryptic properties of a cluster of dominant flavivirus cross-reactive antigenic sites. *J. Virol.* **80**, 9557–9568 (2006).
- Vogt, M. R. *et al.* Poorly neutralizing cross-reactive antibodies against the fusion loop of West Nile virus envelope protein protect *in vivo* via Fc γ receptor and complement-dependent effector mechanisms. *J. Virol.* **85**, 11567–11580 (2011).
- Balsitis, S. J. *et al.* Lethal antibody enhancement of dengue disease in mice is prevented by Fc modification. *PLoS Pathog.* **6**, e1000790 (2010).
- Dejinirattisai, W. *et al.* Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with Zika virus. *Nat. Immunol.* <http://dx.doi.org/10.1038/ni.3515> (2016).
- Dejinirattisai, W. *et al.* A new class of highly potent, broadly neutralizing antibodies isolated from viremic patients infected with dengue virus. *Nat. Immunol.* **16**, 170–177 (2015).
- Goncalves, A. P., Engle, R. E., St Claire, M., Purcell, R. H. & Lai, C. J. Monoclonal antibody-mediated enhancement of dengue virus infection *in vitro* and *in vivo* and strategies for prevention. *Proc. Natl Acad. Sci. USA* **104**, 9422–9427 (2007).
- Halstead, S. B. *In vivo* enhancement of dengue virus infection in rhesus monkeys by passively transferred antibody. *J. Infect. Dis.* **140**, 527–533 (1979).
- Rouviniski, A. *et al.* Recognition determinants of broadly neutralizing human antibodies against dengue viruses. *Nature* **520**, 109–113 (2015).
- Dai, L. *et al.* Structures of the Zika virus envelope protein and its complex with a flavivirus broadly protective antibody. *Cell Host Microbe* **19**, 696–704 (2016).
- Kuhn, R. J., Dowd, K. A., Beth Post, C. & Pierson, T. C. Shake, rattle, and roll: Impact of the dynamics of flavivirus particles on their interactions with the host. *Virology* **479–480**, 508–517 (2015).
- Sabchareon, A. *et al.* Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. *Lancet* **380**, 1559–1567 (2013).
- Vannice, K. S., Durbin, A. & Hombach, J. Status of vaccine research and development of vaccines for dengue. *Vaccine* **34**, 2934–2938 (2016).
- Plevka, P. *et al.* Maturation of flaviviruses starts from one or more icosahedrally independent nucleation centres. *EMBO Rep.* **12**, 602–606 (2011).
- Cherrier, M. V. *et al.* Structural basis for the preferential recognition of immature flaviviruses by a fusion-loop antibody. *EMBO J.* **28**, 3269–3276 (2009).
- Beltramello, M. *et al.* The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. *Cell Host Microbe* **8**, 271–283 (2010).
- Dowd, K. A., Mukherjee, S., Kuhn, R. J. & Pierson, T. C. Combined effects of the structural heterogeneity and dynamics of flaviviruses on antibody recognition. *J. Virol.* **88**, 11726–11737 (2014).
- Lee, P. D. *et al.* The Fc region of an antibody impacts the neutralization of West Nile viruses in different maturation states. *J. Virol.* **87**, 13729–13740 (2013).
- Mukherjee, S. *et al.* Mechanism and significance of cell type-dependent neutralization of flaviviruses. *J. Virol.* **88**, 7210–7220 (2014).
- Capeding, M. R. *et al.* Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. *Lancet* **384**, 1358–1365 (2014).
- Hessell, A. J. *et al.* Fc receptor but not complement binding is important in antibody protection against HIV. *Nature* **449**, 101–104 (2007).

Acknowledgements We thank M. van Kerkhove, the ‘working group ZIKA’ and our colleagues in the Unité de Virologie Structurale at Institut Pasteur for help and discussions; the staff at the crystallography platform and the recombinant proteins platform of the Institut Pasteur; P. Sylvestre and the Risk Prevention Unit of the Institut Pasteur for their biosafety expertise; W. Shepard and P. Legrand for help during data collection and for discussions; the staff at beamlines PX1 and PX2 at SOLEIL synchrotron (Saclay, France), at beamline ID23-2 of the European Synchrotron Radiation Facility (Grenoble, France) and at beamline PX1 of the Swiss Light Source synchrotron (Villigen, Switzerland); G. Bricogne for suggestions on processing anisotropic diffraction data; A. Reiter and H. Prechler for help in recombinant sE production. G.R.S. is a Wellcome Trust Senior Investigator. We acknowledge support from the European Commission FP7 Programme for the DENFREE project under Grant Agreement number 282 378FP7 (F.A.R., J.M., G.R.S. and A.Sa.); the “Integrative Biology of Emerging Infectious Diseases” Labex (Laboratoire d’Excellence) grant number ANR-10-LABX-62-IBED (French Government’s “Investissements d’Avenir” program) (F.A.R.); the transnational ANR/FWF grant FlaviStem/11378 (F.A.R. and K.S.), the Medical Research Council, UK (J.M.); the National Institute for Health Research Biomedical Research Centre, Funding Scheme, UK (G.R.S.); and the NEUTRAVIR grant from Région Ile-de-France (DIM-Maladies Infectieuses) (P.E., A.H. and F.A.R.).

Author Contributions The first four authors (G.B.-S., W.D., A.R. and M.-C.V.) contributed equally to this work and are listed in alphabetical order. G.B.-S. did the first neutralization experiments and contributed throughout the project. V.-M.C.-L., A.Sa. and E.S.L. provided ZIKV from both PF13 and HD78788 strains. I.M., K.S. and F.X.H. provided the recombinant sE protein. A.R. and A.Sh. prepared recombinant antibody fragments; A.R., G.B.-S. and W.D. did neutralization and binding experiments; A.R. grew the crystals and collected synchrotron data together with A.H.; M.-C.V. collected synchrotron data, processed the data, built, refined and analysed the atomic models. E.S.-L. did the phylogenetic analyses; A.R., A.Sh. and P.E. did the biolayer interferometry experiments with recombinant sE. A.R., G.B.-S., K.S., J.M., F.X.H., G.R.S. and F.A.R. planned the experiments. F.A.R. wrote the paper.

Author Information Atomic coordinates and structure factors amplitudes have been deposited in the Protein Data Bank (PDB) under accession numbers 5LBS (for complex ZIKV sE-EDE1 C8 scFv), 5LCV (for complex ZIKV sE-EDE2) and 5LVB (for A11 Fab and unliganded ZIKV sE). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.A.R. (felix.rey@pasteur.fr), G.R.S. (g.screaton@imperial.ac.uk) or F.X.H. (franz.x.heinz@meduniwien.ac.at).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Recombinant production of ZIKV sE protein. Recombinant Zika virus sE protein (strain H/PF/2013, GenBank accession number KJ776791) was produced with a tandem strep-tag in the *Drosophila* Expression System (Invitrogen) as described previously^{29,30}. A chemically synthesized DNA fragment (GeneArt) containing the Zika sE sequence (amino acids 1–408) was cloned into the expression vector pT389 (ref. 31) that encodes the export signal sequence BIP, an enterokinase cleavage site and the strep-tag. *Drosophila* Schneider 2 (S2) cells were stably transfected using blasticidin for selection. Protein expression was induced by the addition of CuSO₄ and supernatants were collected 7–10 days after induction. Antigens were purified via affinity chromatography with Streptactin columns (IBA) according to the manufacturer's instructions. A final purification gel filtration step used a Superdex increase 200 10/300 GL column equilibrated in 50 mM Tris, pH 8, 500 mM NaCl.

Production of antigen-binding (Fab) and scFv fragments of the bnAbs. The bnAb fragments were cloned into plasmids for expression as Fab³² and scFv³³ in *Drosophila* S2 cells. The constructs contain a tandem strep-tag fused at the C terminus (only of the heavy chain in the case of the Fab) for affinity purification. The purification protocol included a Streptactin affinity column followed by gel filtration as described above.

Expression of human monoclonal anti-DENV E antibodies. Full IgG antibodies were produced in 293T cells (a gift from C. Lee), which were free from mycoplasma contamination tested by Lookout Mycoplasma PCR detection kit (Sigma MP0035). These cells were co-transfected with plasmids containing heavy and light chains of immunoglobulin G1 as described previously¹³.

Immune complex formation and isolation. The purified ZIKV sE protein was mixed with Fab A11 or scFv C8 (in approximately twofold molar excess) in standard buffer (500 mM NaCl, Tris 50 mM, pH 8.0). The volume was brought to 0.5 ml by centrifugation in a Vivaspin 10 kDa cut-off; after a 30-min incubation at 4 °C, the complex was separated from excess Fab or scFv by size-exclusion chromatography for ZIKV sE and scFv C8. For ZIKV sE and Fab A11, no apparent complex formation could be seen in size-exclusion chromatography; therefore a solution containing sE at a concentration of 1.5 mg ml⁻¹ and Fab A11 at a concentration of 3 mg ml⁻¹ (corresponding to a molar ratio ~1:2 antigen:antibody) was directly used for crystallization. In all cases, the buffer was exchanged to 150 mM NaCl, 15 mM Tris, pH 8, for crystallization trials. The protein concentrations used for crystallization, determined by measuring the absorbance at 280 nm and using an extinction coefficient estimated from the amino acid sequences, are listed in Extended Data Table 1.

Real-time biolayer interferometry binding assays. The interactions of purified ZIKV E protein with IgG FLE P6B10, IgG EDE1 C8, IgG EDE2 A11, and control IgG 28C (an anti-influenza virus) were monitored in real-time using a bio-layer interferometry Octet-Red384 device (Pall ForteBio). Anti-human IgG Fc capture biosensors (Pall ForteBio) were loaded for 10 min at 1,000 r.p.m. shaking speed using antibodies at 5 µg ml⁻¹ in assay buffer (PBS plus 0.2 mg ml⁻¹ BSA and tween 0.01%). Unbound antibodies were washed away for 1 min in assay buffer. IgG-loaded sensors were then incubated for 15 min at 1,200 r.p.m. in the absence and presence of twofold serially diluted ZIKV sE protein concentrations in assay buffer. Molar concentrations were calculated for the sE protein in a dimeric form. For the antibodies FLE P6B10, EDE1 C8 and EDE2 A11, the following ZIKV sE concentration ranges: 0.78–50 nM, 3.125–200 nM and 50–3,200 nM, were respectively used. Reference binding experiments were carried out in parallel on sensors loaded with control IgG 28C. Dissociation of the complexes formed was then monitored for 10 min by dipping sensors in assay buffer alone. Operating temperature was maintained at 25 °C. The real-time data were analysed using Scrubber 2.0 (Biologic Software) and Biaevaluation 4.1 (GE Healthcare). Specific signals were obtained by double-referencing, that is, subtracting non-specific signals measured on non-specific IgG-loaded sensors and buffer signals on specific IgG-loaded sensors. Association and dissociation profiles, as well as steady-state signal versus concentration curves, were fitted assuming a 1:1 binding model.

Crystallization and X-ray structure determinations. The crystallization and cryo-cooling conditions for diffraction data collection are listed in Extended Data Table 1. Crystallization trials were performed in sitting drops of 400 nl. Drops were formed by mixing equal volumes of the protein and reservoir solution in 96-well Greiner plates, using a Mosquito robot and monitored by a Rock-Imager. Crystals were optimized using a robotized Matrix Maker and Mosquito

setups on 400 nl sitting or hanging drops, or manually in 24-well plates using 2–3 µl hanging drops.

Because of the strong anisotropy of the crystals (see results for anisotropy in Extended Data Table 1), an important number of crystals was tested at several beam lines at different synchrotrons (SOLEIL, St Aubin, France; ESRF, Grenoble, France; SLS, Villigen, Switzerland). The crystals having the less anisotropic diffraction data were used to determine the structures. The data sets were indexed, integrated, scaled and merged using XDS³⁴ and AIMLESS³⁵. A preliminary model of ZIKV sE protein was built from the DENV-2 sE (4UTA) structure using the structure homology-modelling server SWISS-MODEL³⁶. The structures of the complexes were then determined by molecular replacement with PHASER³⁷ using the search models listed in Extended Data Table 1. AIMLESS and PHASER programs were used within the CCP4 suite³⁸.

The DEBYE and STARANISO programs developed by Global Phasing Ltd were applied to the data scaled with AIMLESS without applying a resolution limit, using the STARANISO server (<http://staraniso.globalphasing.org/>). These programs perform an anisotropic cut-off of merged intensity data on the basis of an analysis of local $I/\sigma(I)$, compute Bayesian estimates of structure amplitudes, taking into account their anisotropic fall-off, and apply an anisotropic correction to the data. These corrected anisotropic amplitudes were then used for further refinement of the structures with BUSTER/TNT³⁹. Please note that the Extended Data Table 1 shows the refinement statistics against the full sets of reflections truncated at the best high-resolution along the h , k or l axis.

The models were then alternatively manually corrected and completed using COOT⁴⁰ and refined using BUSTER/TNT against the amplitudes corrected for anisotropy. Refinements were constrained using non-crystallographic symmetry. The refined structures have the following final $R_{\text{work}}/R_{\text{free}}$ (in %) values: ZIKV sE–EDE1 C8 scFv (19.5/22.1), ZIKV sE–EDE2 A11 Fab (22.3/23.7) and unliganded ZIKV sE (20.8/23.6) (see Extended Data Table 1).

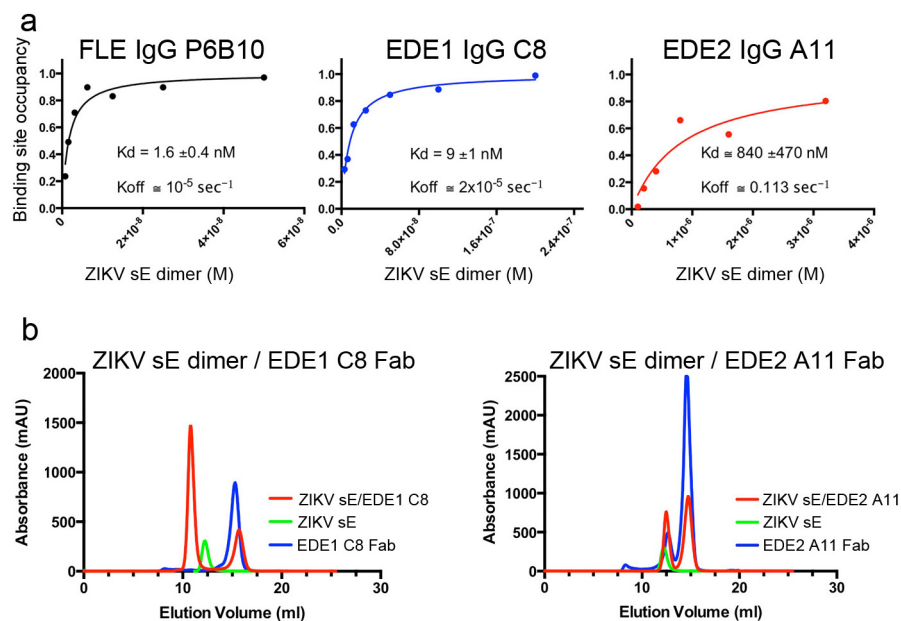
Analysis of the atomic models and illustrations. Each complex was analysed with the CCP4 suite of programs and the polar contacts were computed with the PISA website⁴¹. For the intermolecular interactions shown in Extended Data Figs 2 and 3 and Extended Data Tables 4 and 5, the maximal cut-off distances used were 4 Å and 4.75 Å for polar and van der Waals contacts, respectively. Multiple sequence alignments were calculated using Clustal W and Clustal X version 2 (ref. 42) on the EBI server⁴³. The figures illustrating the structural models were prepared using ESPript⁴⁴ and the PyMOL Molecular Graphics System, version 1.5.0.4 (Schrödinger) (<http://pymol.sourceforge.net>).

Phylogenetic trees. The maximum likelihood phylogenetic trees were inferred using 12 representative amino acid sequences of Flavivirus envelope protein E or RNA polymerase NS5 proteins, using the LG model available in PhyML⁴⁵ and a combination of SPR+NNI branch-swapping. Bootstrap values were calculated from 100 bootstrap replicates. The trees were visualized using Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>). The accession codes of sequences used in the tree: Zika virus (ZIKV, KJ776791, strain H-PF-2013_French_Polynesia); dengue virus serotype 1 (DENV-1, NC_001477); dengue virus serotype 2 (DENV-2, NC_001474); dengue virus serotype 3 (DENV-3, NC_001475); dengue virus serotype 4 (DENV-4, NC_002640); Saint Louis encephalitis virus (SLEV, NC_007580); Japanese encephalitis virus (JEV, NC_001437); Murray Valley encephalitis virus (MVEV, NC_000943); West Nile virus (WNV, NC_001563); yellow fever virus (YFV, NC_002031); tick-borne encephalitis virus (TBEV, NC_001672); and Powassan virus (POWV, NC_003687).

Viral stocks. The African strain Zika HD78788 was obtained from the Institut Pasteur collection and the Asian strain Zika PF13, isolated from a patient during ZIKV outbreak in French Polynesia in 2013, was obtained through the DENFREE (FP7/2007–2013) consortium. Viral stocks were prepared from supernatant of infected C6/36 cells (ATCC CRL-1660) clarified by centrifugation at 3,000g at 4 °C and titrated on Vero cells (ATCC CRL-1586) by a focus-forming assay. Stocks were kept at –80 °C until use. All cell lines were free from mycoplasma contamination.

Neutralization assays. Virus neutralization by the tested human antibodies was evaluated using a focus reduction neutralization test (FRNT). About 100 focus-forming units from virus stocks were incubated with a serial dilution of antibody for 1 h at 37 °C. The mixture was then added to Vero cells and foci were left to develop in presence of 1.5% methylcellulose for 2 days. Foci were then stained after fixation with 4% formaldehyde using anti-E 4G2 antibody (ATCC HB-112) and anti-mouse horseradish peroxidase (HRP)-conjugated secondary antibody (ThermoFisher 31430). The foci were visualized by diaminobenzidine (DAB) (Sigma D5905) staining and plates were counted using the ImmunoSpot S6 Analyser (Cellular Technology Limited, CTL). Neutralization curves and 50% FRNT values were calculated by nonlinear regression analysis using Prism 6, GraphPad software.

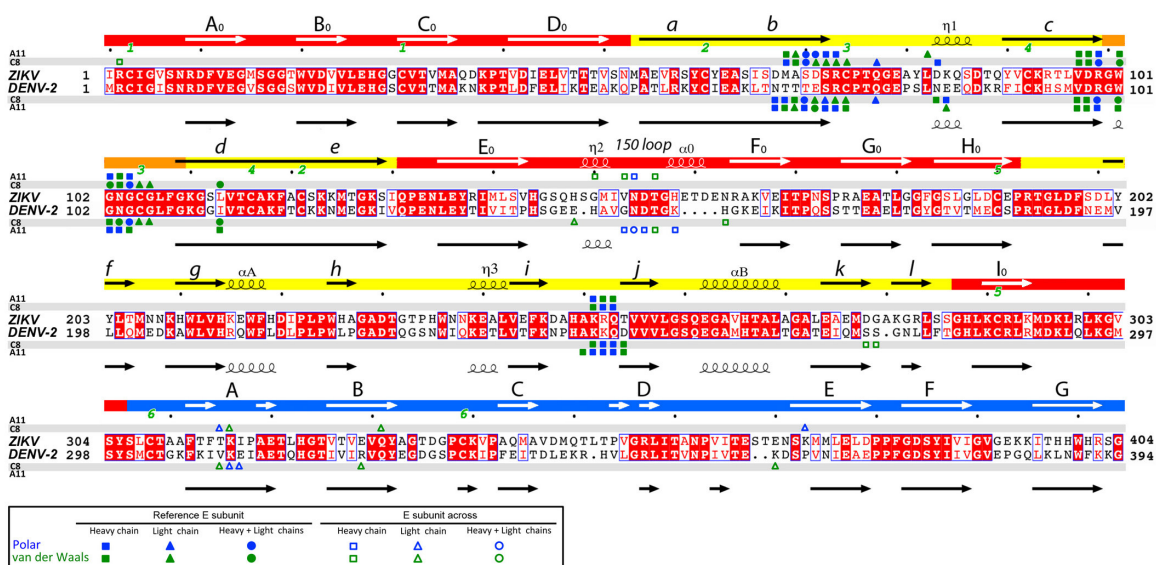
29. Vratskikh, O. *et al.* Dissection of antibody specificities induced by yellow fever vaccination. *PLoS Pathog.* **9**, e1003458 (2013).
30. Jarmer, J. *et al.* Variation of the specificity of the human antibody responses after tick-borne encephalitis virus infection and vaccination. *J. Virol.* **88**, 13845–13857 (2014).
31. DuBois, R. M. *et al.* Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* **493**, 552–556 (2013).
32. Backovic, M. *et al.* Efficient method for production of high yields of Fab fragments in *Drosophila* S2 cells. *Protein Eng. Des. Sel.* **23**, 169–174 (2010).
33. Gilmartin, A. A. *et al.* High-level secretion of recombinant monomeric murine and human single-chain Fv antibodies from *Drosophila* S2 cells. *Protein Eng. Des. Sel.* **25**, 59–66 (2012).
34. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
35. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
36. Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–W258 (2014).
37. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
38. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
39. Bricogne, G. *et al.* BUSTER version 2.10.2 (Global Phasing Ltd, 2016).
40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
41. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
42. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
43. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
44. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
45. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).



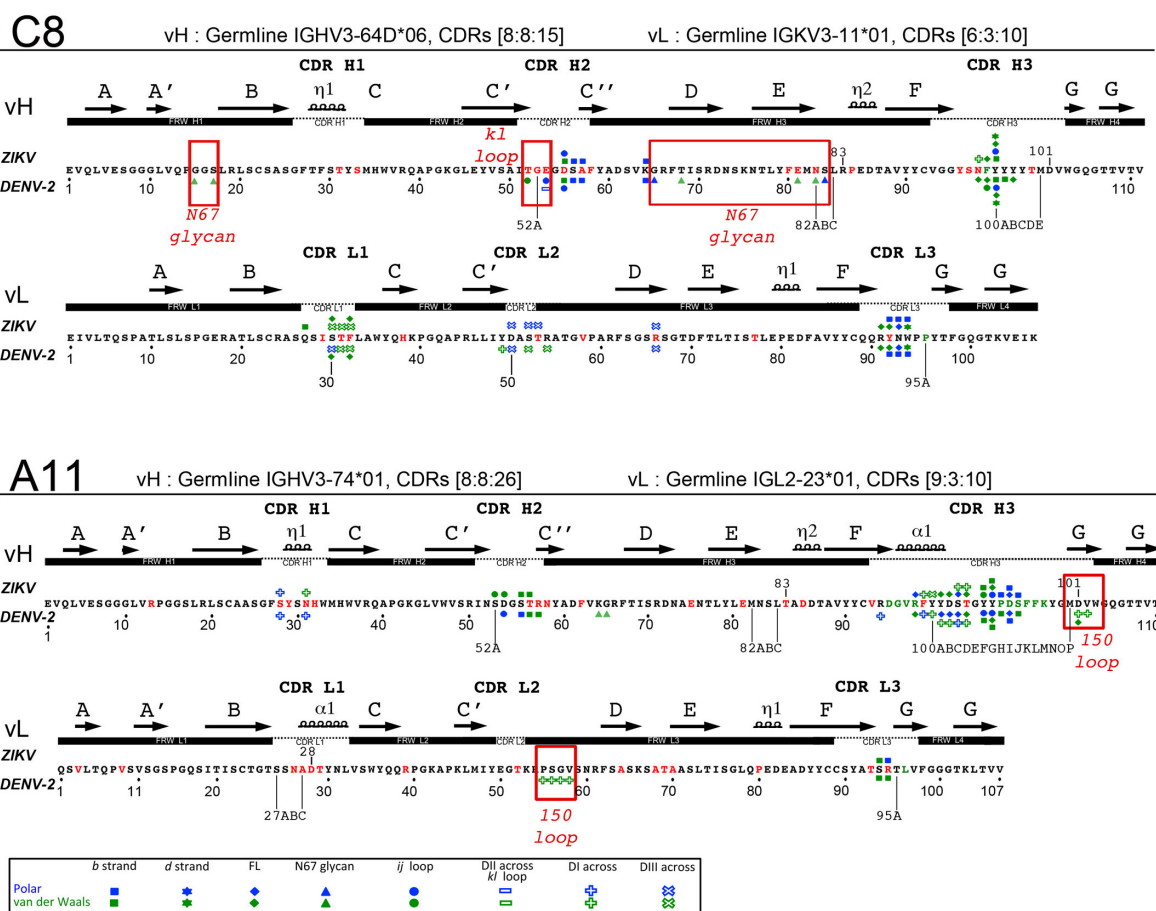
Extended Data Figure 1 | Antibody binding to recombinant ZIKV protein. a, Biolayer interferometry experiments plotted on a linear scale. The antibodies were immobilized on the biosensor tip, and the ZIKV sE protein was in solution at the indicated concentrations. The antibody used is indicated in each plot. Note that the horizontal scale is different

for the three antibodies. The estimated dissociation constant (K_d) and estimated dissociation rate (K_{off}) values are indicated. **b,** Size-exclusion chromatography results for isolated sE, isolated Fab fragments, and ZIKV sE plus Fab fragments, as indicated.

a



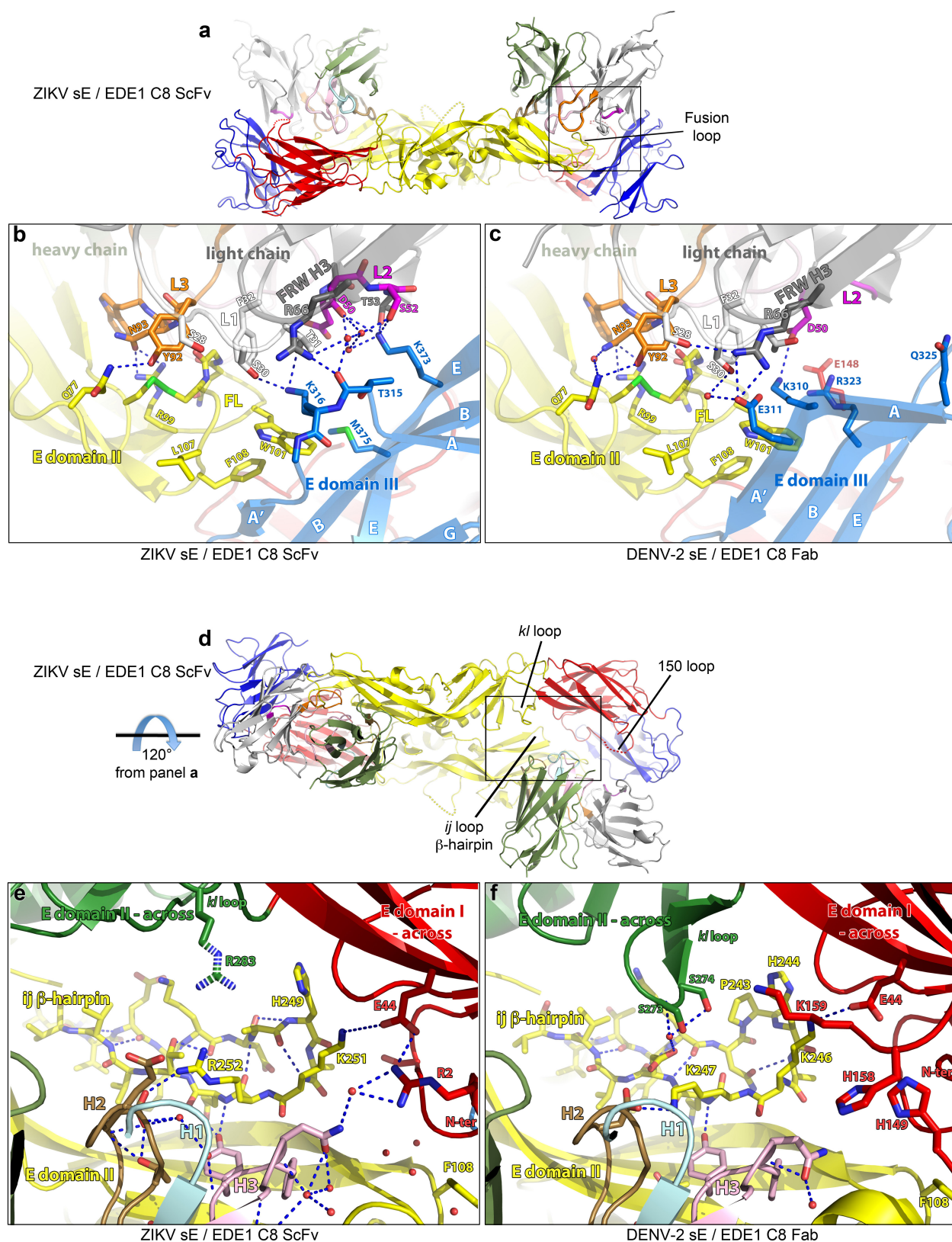
b



Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Residues involved in bnAb–antigen interactions. **a**, Antibody contacts on the amino acid sequence alignment of ZIKV and DENV-2 sE. A red background highlights identical residues. Secondary structure elements are indicated together with their labels above (ZIKV) and below (DENV-2) the sequences. The domain organization of ZIKV and DENV-2 sE is symbolized by a coloured bar above the sequences (domain I red, domain II yellow, domain III blue and the fusion loop orange). Residues involved in polar and van der Waals protein–protein contacts are marked using blue and green symbols, respectively, as indicated in the inset key, displayed above and below the alignment for ZIKV and DENV-2 sE, respectively. Full and empty symbols correspond to antibody contacts on the reference subunit of sE (defined as the one contributing the fusion loop to the epitope) and the opposite subunit of sE, respectively. Residues contacted only by the heavy or light chain are marked with squares or triangles, respectively, and those contacted by both antibody chains with circles. Dots above the sequences mark every 10 residues on the ZIKV sE sequence. Disulfide bridges are numbered in green above the sequences. **b**, Amino acid sequence of the

heavy and light chains variable domains (vH and vL) of bnAbs EDE1 C8 (top) and EDE2 A11 (bottom) with the framework (FRW) indicated by black bars and IMGT CDR regions by thin dashed lines. The secondary structure elements of the Ig vH and vL β -barrels are indicated above the sequences. Somatic mutations are in red and residues arising from recombination at the V-D-J junction are in green. Symbols above and below the sequences mark residues involved in contacts with ZIKV and DENV-2 sE, respectively, coded for the contacted site in sE as indicated in the key (inset at the bottom). Polar and van der Waals contacts are shown in blue and green, respectively. The antibody residues contacting the reference sE subunit (defined as the one contributing the fusion loop to the epitope) are marked by plain colour symbols while those making contact across the dimer interface by empty coloured symbols. Red boxes highlight the contacts found in the DENV-2 sE complex and absent in the ZIKV sE complex, involving N67 glycan, *kl* and 150 loops. The details of the polar contacts are listed in the Extended Data Tables 4 and 5 (see also Fig. 3e, f). The predicted vH and vL germline alleles are indicated with the corresponding CDR lengths (see Table 1 in ref. 16).



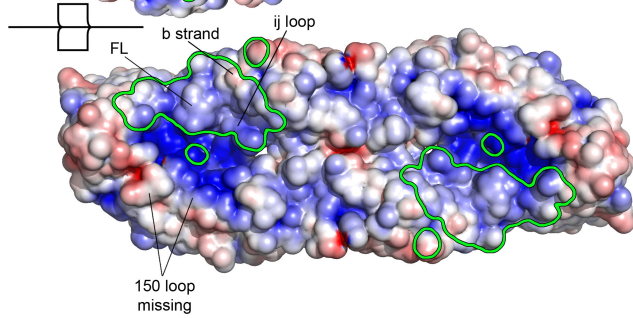
Extended Data Figure 3 | Details of EDE1 C8 bnAb contact across the dimer interface. **a**, Overall view of the ZIKV sE-EDE1 C8 scFv complex. The box indicates the region zoomed in **b**. **b**, Details of the interactions of the C8 light chain with domain III across the dimer interface. **c**, Same region for the DENV-2 sE-EDE1 C8 Fab complex. Note that the sE residues involved are different. **d**, The complex rotated by 120° (as indicated by the arrow) to show the interaction in the *ij* loop, enlarged in **e**. **e**, The *ij* loop is displayed in sticks, to show the interaction of its main

chain with the antibody. Domain II from the subunit across is coloured green to distinguish from domain II of the reference subunit; the dashed sticks for the arginine residue indicate that it has poor electron density in the crystal. **f**, Same view of the complex with DENV-2. Note that the residues from across the dimer interface that contact the antibody are different. The residues in the various CDRs are colour coded, matching their label colour (as in Figs 3 and 4).

ZIKV sE / EDE1 C8 ScFv

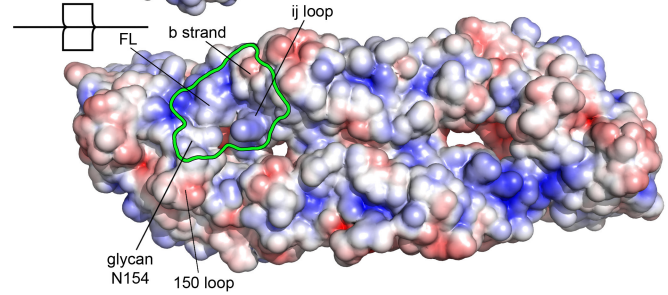
EDE1 C8 ScFv

EDE1 C8 ScFv



ZIKV sE / EDE2 A11 Fab

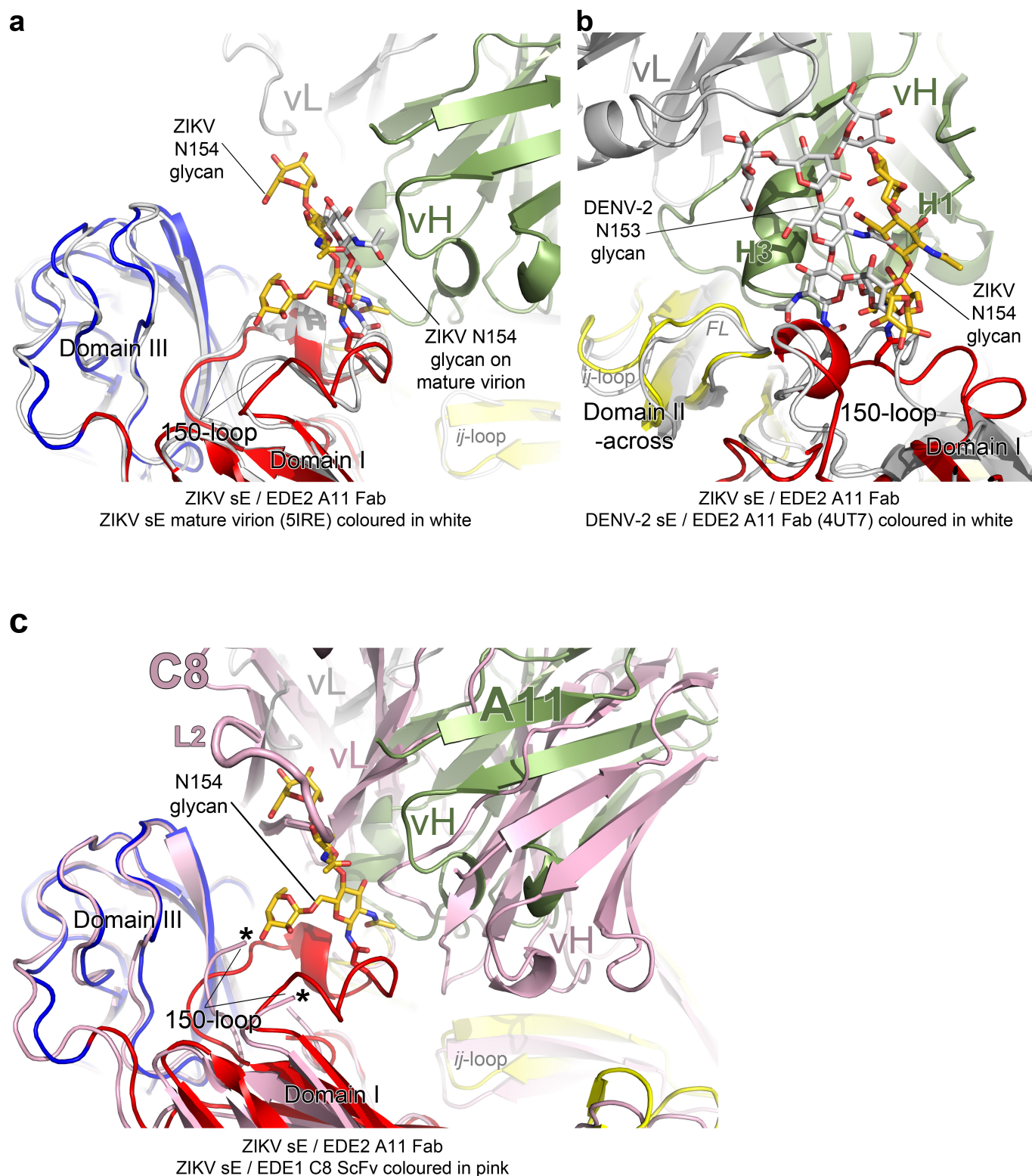
EDE2 A11 Fab



-5.0 kT/e  +5.0 kT/e

Extended Data Figure 4 | Surface electrostatic potential on an open-book representation of the immunocomplexes. The electrostatic potential is coloured according to the bar underneath. The antibody footprints are outlined in green. The disordered 150 loop in the complex

with C8 (left) results in a positive surface patch at one edge of the epitope, which is counteracted by the residues in the 150 loop, as shown on the right in the complex with A11 where this loop is ordered.



Extended Data Figure 5 | Details of the A11 interaction with the glycan on the 150 loop. **a**, Superposition of the ZIKV sE-A11 complex (in colours) on the E protein from the cryo-EM structure of the mature virion (ref. 5) (PDB code 5IRE) in white. The E protein was superimposed on the tip of domain II of the reference subunit together with domain III from the opposite subunit. It shows that the 150 loop adopts essentially the same conformation, although fewer sugar residues are visible in the absence of the antibody. **b**, Superposition of the ZIKV sE-A11 complex (in colours) on the DENV-2 sE-A11 complex (in white). The variable domains of the antibody from the two structures were superimposed on each other. Note that in DENV-2 the glycan packs against the α -helix of

the CDR H3, whereas in ZIKV sE the glycan is too far to make the same interaction. **c**, The ZIKV sE-C8 complex (in pink) was superimposed on the ZIKV sE-A11 complex (in colours), to show the clash of the C8 light chain with the glycan, forcing it to move out of the way and be disordered. The superposition also shows that EDE1 C8 reaches further in to contact the *ij* loop and the *kl* loop of the adjacent subunit, as well as domain III. As in **a**, the superposition was done using the tip of domain II of the reference subunit and domain III of the adjacent subunit in the dimer as anchors. The two black asterisks mark the places where the electron density of the 150 loop is lost, resulting in no density in the ZIKV sE-C8 crystal for the short helix, nor for the glycan.

Extended Data Table 1 | Crystallization conditions, data collection and refinement statistics

	ZIKV sE - EDE1 C8 scFv	ZIKV sE - EDE2 A11 Fab	ZIKV sE
Protein Data Bank code	5LBS	5LCV	5LBV
Crystallization conditions			
Protein conc. (mg/ml)*	1.3 (complex)	1.5 (ZIKV sE) 3 (EDE2 A11 Fab)	1.4
Crystallization buffer	1.26 M (NH ₄) ₂ SO ₄ , 0.1 M CHES pH 9.5, 0.2 M NaCl	3.5 M Na Formate, 0.1 M Tris pH 8.5	28.6% PEG 8K 0.1 M Bicine pH 8.3
Crystallization method	sitting drop at 18°C	sitting drop at 18°C	hanging drop at 18°C
Cryo-protectant	20% ethylene glycol in 67% of crystallization solution	16% glycerol in 67% of crystallization solution	oil mixture paraffin:paratone (1:1)
Data Collection^{†,§}			
Beamline	SOLEIL, Proxima 1	SOLEIL, Proxima 2	SOLEIL, Proxima 1
Detector	Pilatus 6M	Eiger 9M	Pilatus 6M
Space group	P 2 ₁ 2 ₁	C 2 2 2 ₁	C 2 2 2 ₁
Unit cell: a, b, c (Å)	60.8, 121.3, 257.8	204.3, 207.3, 124.6	64.6, 213.6, 124.0
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	40.0- 2.41 (2.46-2.41)	40.0- 2.64 (2.69-2.64)	38.5- 2.2 (2.25-2.20)
Anisotropy direction[†]			
Resolution where CC _{1/2} > 0.50			
overall (Å)	2.69	2.87	2.2
along h axis (Å)	3.24	4.23	3.26
along k axis (Å)	2.95	2.64	2.2
along l axis (Å)	2.41	2.9	2.2
Measured reflections	560 284 (35 314)	760 128 (38 976)	318 556 (18 018)
Unique reflections	74 842 (4 571)	77 483 (4 547)	44 168 (3 025)
Completeness (%)	100 (100)	99.8 (99.4)	99.5 (93.4)
Mn(I) half-set correlation	0.99 (0.15)	0.98 (0.17)	0.99 (0.33)
Mean I/σ(I)	6.7 (1.4)	5.1 (0.4)	9.4 (0.5)
Multiplicity	7.5 (7.7)	9.8 (8.6)	7.2 (6.0)
B wilson (Å ²)	41.2	47.4	60.1
Rmerge	0.33 (3.1)	0.50 (6.6)	0.10 (3.3)
Rmeas	0.36 (2.9)	0.53 (7.1)	0.11 (3.7)
Rpim	0.13 (1.2)	0.17 (2.4)	0.04 (1.5)
Structure Determination			
MR search models	ZIKV sE from ZIKV sE - EDE2 A11 Fab complex (5LCV) EDE1 C8 Fab variable domain (4UTA)	ZIKV model sE from DENV-2 sE (4UTA) EDE2 A11 scFv (4UT7) EDE2 A11 Fab constant domain (4UTB)	ZIKV sE from ZIKV sE - EDE1 C8 scFv complex (5LBS)
NCS restraint	2	2 applied only on sE dimer	2
Targeting	NA	EDE2 A11 scFv (4UT7)	NA
Number of TLS groups	12	12	12
Refinement[§]			
Resolution cut-off (Å)	40.0- 2.41	40.0- 2.64	40.0- 2.2
Rwork (%) / Rfree (%)	19.5 / 22.1	22.3 / 23.7	20.8 / 23.6
N° of Work / Free reflections	74 785 / 3 719	76 253 / 3 840	44 038 / 2 219
 atomic factors (Å) ²	72.6	89.3	64.1
N° of protein atoms	9 595	9 495	5 978
N° of heteroatoms	212	43	43
R.m.s.d. from ideal			
Bond lengths (Å)	0.01	0.01	0.01
Bond angles (°)	1.22	1.28	1.24
Ramachandran			
Favoured (%)	96.5	95.7	97.3
Allowed (%)	3.26	3.65	1.86
Outliers (%)	0.24	0.65	0.84

The ZIKV sE buffer used for all crystallization experiments was: 150 mM NaCl, 15 mM Tris, pH 8.

*The protein concentration was estimated using theoretical extinction coefficients of the complexes (ZIKV sE + Fab or scFv). Absorbance at 280 nm ($A_{280\text{ nm}}$) of the protein solution was measured before crystallization. The theoretical extinction coefficients for individual component are: ZIKV sE: 1.345; bnAb EDE2 A11 Fab: 1.68 (see Methods for more details); bnAb EDE1 C8 scFv: 1.9. Extinction coefficients were calculated without taking carbohydrate moieties into account. One crystal was used to collect a diffraction data set for each complex to determine the structure.

†The anisotropy statistics were computed with AIMLESS.

‡Highest resolution shell is shown in parentheses.

§Low-resolution for data processing and refinements was truncated to 40 Å.

||Ramachandran statistics were calculated with MolProbity.

CC1/2, correlation coefficient; *h*, *k* and *l*: indices that define the lattice planes; *I*/σ(*I*), empirical signal-to-noise ratio; MR, molecular replacement; NCS, non-crystallographic symmetry; Rmeas, multiplicity-corrected R; Rpim, expected precision; TLS, parameterization describing translation, liberation and screw-motion to model anisotropic displacements.

Extended Data Table 2 | The r.m.s.d. values between sE dimers in the various structures of ZIKV and DENV-2

ZIKV*	PDB code	ZIKV sE (this work)	ZIKV sE-EDE2 A11 Fab (this work)	ZIKV sE-EDE1 C8 scFv [†] (this work)	ZIKV sE (Dai <i>et al.</i> , 2016)	ZIKV mature virion sE dimer from au	ZIKV mature virion icosahedral sE dimer	ZIKV thermally stable mature virion sE dimer from au	ZIKV thermally stable mature virion icosahedral sE dimer
		5LBV	5LCV	5LBS	5JHM	5IRE	5IRE	5IZ7	5IZ7
ZIKV sE (this work)	5LBV	/							
ZIKV sE-EDE2 A11 Fab	5LCV	3.33 (796)	/						
ZIKV sE-EDE1 C8 scFv [†]	5LBS	2.31 (776) 2.29 (764)	2.41 (773) 2.29 (761)	0.57 (764) [‡]					
ZIKV sE (Dai <i>et al.</i> , 2016)	5JHM	1.8 (769)	4.1 (768)	3.06 (771) 2.93 (761)	/				
ZIKV mature virion sE dimer from au	5IRE	3.22 (800)	1.72 (799)	1.57 (778) 1.69 (764)	3.72 (775)	/			
ZIKV mature virion icosahedral sE dimer	5IRE	2.88 (800)	1.76 (799)	1.60 (778) 1.71 (764)	3.74 (775)	0.46 (806) [§] 0.45 (1002)	/		
ZIKV thermally stable mature virion sE dimer from au	5IZ7	3.03 (800)	1.80 (799)	1.75 (778) 1.88 (764)	3.89 (775)	1.36 (806) [§] 1.57 (1002)	1.36 (806) [§] 1.56 (1002)	/	
ZIKV thermally stable mature virion icosahedral sE dimer	5IZ7	2.94 (800)	1.84 (799)	1.67 (778) 1.80 (764)	3.78 (775)	1.36 (806) [§] 1.56 (1002)	1.36 (806) [§] 1.54 (1002)	1.41 (806) [§] 1.43 (1008)	/

ZIKV/DENV-2*	PDB code	ZIKV sE (this work)	ZIKV sE-EDE2 A11 Fab (this work)	ZIKV sE-EDE1 C8 scFv [†] (this work)	ZIKV sE (Dai <i>et al.</i> , 2016)	ZIKV sE mature virion sE dimer from au	ZIKV sE mature virion icosahedral sE dimer	ZIKV thermally stable mature virion sE dimer from au	ZIKV thermally stable mature virion icosahedral sE dimer
		5LBV	5LCV	5LBS	5JHM	5IRE	5IRE	5IZ7	5IZ7
DENV-2 sE	4UTC	4.57 (766)	6.47 (768)	5.49 (756) 5.55 (742)	4.75 (753)	5.94 (772)	5.94 (772)	6.23 (757)	6.18 (757)
DENV-2 sE-EDE2 A11 Fab	4UTB	4.45 (773)	7.08 (774) 0.29/0.28 (241) [¶]	5.61 (763) 5.57 (744)	4.2 (754)	6.29 (773)	6.28 (773)	7.38 (763)	7.31 (763)
DENV-2 sE-EDE1 C8 Fab	4UTA	6.04 (745)	8.17 (744)	6.88 (748) 6.78 (736)	5.71 (747)	7.52 (751)	7.49 (751)	7.65 (738)	7.57 (738)
DENV-2 mature virion sE dimer from au	3J27	3.89 (779)	2.26 (778)	2.36 (769) 2.81 (755)	4.82 (766)	2.01 (785) [§] 2.09 (979)	2.04 (785) [§] 2.13 (979)	4.33 (770) [§] 2.15 (988)	4.36 (770) [§] 2.20 (988)
DENV-2 mature virion icosahedral sE dimer	3J27	3.59 (779)	2.45 (778)	2.25 (769) 2.68 (755)	4.49 (766)	1.99 (785) [§] 2.07 (979)	1.99 (785) [§] 2.07 (979)	4.31 (770) [§] 2.14 (988)	4.33 (770) [§] 2.15 (988)

DENV-2*	PDB code	DENV-2 sE	DENV-2 sE-EDE2 A11 Fab	DENV-2 sE-EDE1 C8 Fab	DENV-2 mature virion sE dimer from au	DENV-2 mature virion icosahedral sE dimer
		4UTC	4UTB	4UTA	3J27	3J27
DENV-2 sE	4UTC	/				
DENV-2 sE-EDE2 A11 Fab	4UTB	2.34 (775)	/			
DENV-2 sE-EDE1 C8 Fab	4UTA	2.93 (747)	1.53 (741)	/		
DENV-2 mature virion sE dimer from au	3J27	5.26 (779)	6.78 (775)	6.99 (754)	/	
DENV-2 mature virion icosahedral sE dimer	3J27	4.92 (779)	6.37 (775)	6.56 (754)	0.89 (790) [§] 0.88 (990)	/

r.m.s.d. (in Å) computed with PyMOL software using the C α atoms of the sE dimers. In parentheses, the number of C α atoms used for the calculation. au, asymmetric unit.

*The r.m.s.d. is computed using residues 1 to 403 for ZIKV sE, or residues 1 to 395 for DENV-2 sE, except when indicated^{§,||}.

[†]There are two independent half dimers (ZIKV sE-EDE1 C8 scFv) in the asymmetric unit. The r.m.s.d. is computed between the two dimers of sE generated by crystallographic symmetry for each sE in the asymmetric unit.

[‡]This r.m.s.d. is computed between the two independent half dimers of sE in the asymmetric unit.

[§]This r.m.s.d. is computed between the sE dimers excluding stem and transmembrane regions of ZIKV (residues 1 to 403) and DENV-2 (residues 1 to 395).

^{||}This r.m.s.d. is computed between the sE dimers including stem and transmembrane regions of ZIKV and DENV-2 (residues 1 to C-terminal).

[¶]The two r.m.s.d. values refer to the superposition of the variable domains of Fab A11 in ZIKV sE-EDE2 A11 Fab complex on each Fab A11 in DENV-2 sE-EDE2 A11 Fab complex.

Extended Data Table 3 | Buried surface areas and surface complementarity in the ZIKV sE dimer–EDE complexes and in DENV-2 sE dimer–EDE complexes

	BSA BNA Fab or scFv			BSA sE dimer					Complex	
	vH	vL	Total	Reference subunit (glycans*)	Opposite subunit (glycans*)	Total	Main chain atoms†	Total glycan BSA	BSA / molecule	SC
ZIKV sE - EDE1 C8 scFv fragment§										
ZIKV sE dimer Epitope dimer-1	426.1	494.8	920.9	653.0 (NA)	237.2 (NA)	890.2	346.9 (39%)	NA	905.5	0.683
ZIKV sE dimer Epitope dimer-2	438.4	492.4	930.8	678.1 (NA)	223.2 (NA)	901.3	343.9 (38%)	NA	916.0	0.738
DENV-2 sE - EDE1 C8 Fab fragment (4UTA)										
DENV-2 sE dimer Epitope A	718.9 (516.8)†	471.0 (471.0)†	1189.9	919.1 (192.9)	222.2	1141.3	357.8 (31%)	192.9 (17%)	1165.6	0.693
DENV-2 sE dimer Epitope B	831.9 (570.3)†	528.3 (528.3)†	1360.2	945.5 (230.9)	340.0	1285.5	359.0 (28%)	230.9 (18%)	1322.8	0.687
ZIKV sE - EDE2 A11 Fab fragment 										
ZIKV sE dimer Epitope	718.5	75.7	793.4	570.0 (0.0)	217.4 (134.6)	787.4	253.7 (32.2%)	134.6 (17%)	790.4	0.674
DENV-2 sE - EDE2 A11 Fab fragment (4UTB)										
DENV-2 sE dimer Epitope A	923.9 (251.3)†	189.2 (148.8)†	1113.0	531.8 (14.2)	472.6 (342.4)	1004.4	221.5 (22%)	356.6 (35%)	1058.7	0.706
DENV-2 sE dimer Epitope B	954.1 (302.2)†	185.1 (136.7)†	1138.8	460.3 (58.5)	571.3 (341.4)	1031.6	219.3 (21.2%)	400.0 (39%)	1085.2	0.668

BSA, buried surface area (in Å²) of sE protein by the Fabs or scFv (calculated with the program 'areaimol' in CCP4). BSA/molecule, average buried surface area between one Fab or one scFv and the sE dimer. SC, shape complementarity coefficient (calculated with the program 'sc' in CCP4). The dots density used to compute both BSA and SC was set to 15 dots per Å². The van der Waals probe radius was set to 1.4 Å. NA, non applicable.

*In parentheses, contribution of glycan chains to buried surface area: N154 (for ZIKV sE) and N67 and/or N153 (for DENV-2 sE).

†In parentheses, the BSA was computed removing the glycan chains N67 for DENV-2 sE, to compare with the BSA of ZIKV sE, which does not carry the N67 glycan.

‡Contribution of main-chain atoms to BSA. In parentheses, contribution indicated as percentages.

§There are two independent half dimers (ZIKV sE–EDE1 C8 scFv) in the asymmetric unit. The two dimers of ZIKV sE (dimer-1 and dimer-2) are generated by applying crystallographic symmetry for each sE in the asymmetric unit.

||Only one Fab A11 binds to the sE dimer in the ZIKV sE–EDE2 A11 Fab complex.

Extended Data Table 4 | Polar and salt-bridge interactions for ZIKV sE-EDE2 A11 Fab, DENV-2 sE-EDE2 A11 Fab, ZIKV sE-EDE1 C8 scFv and DENV-2 sE-EDE1 C8 Fab

domains of sE	ZIKV sE - EDE2 Fab A11 (5LCV)				DENV-2 sE - EDE2 Fab A11 (4UTB) epitope A				DENV-2 sE - EDE2 Fab A11 (4UTB) epitope B			
	sE	dist (Å)	Fab A11	CDR	sE	dist (Å)	Fab A11	CDR	sE	dist (Å)	Fab A11	CDR
domain I - Reference subunit	b strand	S 70 [O]	2.96	R 95 [NH1]	L3							
		S 70 [O]	2.96	R 95 [NH2]	L3							
		D 71 [OD1]	3.72	R 95 [NH1]	L3							
		D 71 [OD1]	2.66	S 100J [OG]	H3							
		S 72 [OG]	3.21	D 100I [N]	H3							
		S 72 [OG]	2.71	D 100I [OD1]	H3							
		S 72 [O]	3.49	S 100J [N]	H3							
		S 72 [N]	3.39	P 100H [O]	H3							
		R 73 [N]	3.44	D 100I [OD1]	H3							
		R 73 [NE]	2.83	S 100J [OG]	H3							
domain I - across	Fusion loop	R 99 [NH1]	2.78	D 100I [OD1]	H3							
		R 99 [NH1]	3.47	D 100I [OD2]	H3							
		R 99 [NH2]	3.60	D 100I [OD1]	H3							
		R 99 [NH2]	2.70	D 100I [OD2]	H3							
		G 102 [O]	3.16	S 100C [N]	H3							
		G 102 [O]	3.25	S 100C [OG]	H3							
		G 104 [O]	3.78	R 98 [NH1]	H3							
		K 251 [O]	2.72	Y 100G [OH]	H3							
		no glycan at this position										
		P 354 [O]	3.71	N 27B [ND2]	L1							
domain I - across	N153/ N154 glycan	V 153 [O]	3.5	S 100C [OG]	H3							
domain II - reference subunit	b strand	M 68 [O]	3.72	S 5 [OG]	H2							
		M 68 [O]	3.22	A 57 [N]	H2							
		S 70 [OG]	2.94	S 56 [OG]	H2							
		S 70 [N]	3.14	S 56 [OG]	H2							
		S 70 [O]	3.14	W 94 [NE1]	L3							
		S 72 [O]	3.12	N 93 [ND2]	L3							
		Q 77 [NE2]	3.34	Y 92 [OH]	L3							
		D 83 [OD1]	3.56	K 64 [NZ]	FH3							
domain II - reference subunit	Fusion loop	R 99 [NH2]	3.65	N 93 [O]	L3							
		R 99 [NH1]	2.87	N 93 [OD1]	L3							
		R 99 [NH2]	3.02	N 93 [OD1]	L3							
		G 104 [O]	2.74	N 93 [N]	L3							
		R 252 [NH2]	3.37	D 55 [OD2]	H2							
		Q 253 [N]	3.11	Y 100 [OH]	H3							
		Q 253 [O]	3.37	Y 100 [OH]	H3							
		non existing glycan										
		disordered loop										
		disordered glycan										
domain II - reference subunit	ij loop	R 99 [NH2]	3.65	N 93 [O]	L3							
		R 99 [NH1]	2.87	N 93 [OD1]	L3							
		R 99 [NH2]	3.02	N 93 [OD1]	L3							
		G 104 [O]	2.74	N 93 [N]	L3							
		R 252 [NH2]	3.37	D 55 [OD2]	H2							
		Q 253 [N]	3.11	Y 100 [OH]	H3							
		Q 253 [O]	3.37	Y 100 [OH]	H3							
		non existing glycan										
		disordered loop										
		disordered glycan										
domain II - reference subunit	kl loop	disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
		disordered loop										
domain III - across	A strand	T 315 [O]	2.58	R 66 [NH1]	FL3							
domain III - across	E strand	K 373 [NZ]	2.85	S 52 [OG]	L2							
		K 373 [NZ]	2.98	T 53 [OG]	L2							
		K 373 [NZ]	3.56	D 50 [O]	L2							

The polar contacts were computed using the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Bold red font denotes main-chain atoms involved in contacts; bold black font denotes salt bridges; bold blue font denotes glycan interactions. Hydrogen bonds distances cut-off: 3.5 Å; salt-bridge distances cut-off: ≤4 Å. The green background refers to polar contact distances between 3.5 Å and 3.9 Å.

Extended Data Table 5 | Polar and salt-bridge interactions for ZIKV sE-EDE2 A11 Fab, DENV-2 sE-EDE2 A11 Fab, ZIKV sE-EDE1 C8 scFv and DENV-2 sE-EDE1 C8 Fab

[illegible]

The polar contacts were computed with the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Bold red font denotes main-chain atoms involved in contacts; bold black font denotes salt bridges; bold blue font denotes glycan interactions. Hydrogen bonds distances cut-off: 3.5 Å; salt bridge distances cut-off: ≤ 4 Å. The green background refers to polar contact distances between 3.5 Å and 3.9 Å.

Bright carbonate deposits as evidence of aqueous alteration on (1) Ceres

M. C. De Sanctis¹, A. Raponi¹, E. Ammannito^{1,2}, M. Ciarniello¹, M. J. Toplis³, H. Y. McSween⁴, J. C. Castillo-Rogez⁵, B. L. Ehlmann^{5,6}, F. G. Carrozzo¹, S. Marchi^{1,7}, F. Tosi¹, F. Zamboni¹, F. Capaccioni¹, M. T. Capria¹, S. Fonte¹, M. Formisano¹, A. Frigeri¹, M. Giardino¹, A. Longobardo¹, G. Magni¹, E. Palomba¹, L. A. McFadden⁸, C. M. Pieters⁹, R. Jaumann¹⁰, P. Schenk¹¹, R. Mugnuolo¹², C. A. Raymond⁵ & C. T. Russell²

The typically dark surface of the dwarf planet Ceres is punctuated by areas of much higher albedo, most prominently in the Occator crater¹. These small bright areas have been tentatively interpreted as containing a large amount of hydrated magnesium sulfate¹, in contrast to the average surface, which is a mixture of low-albedo materials and magnesium phyllosilicates, ammoniated phyllosilicates and carbonates^{2–4}. Here we report high spatial and spectral resolution near-infrared observations of the bright areas in the Occator crater on Ceres. Spectra of these bright areas are consistent with a large amount of sodium carbonate, constituting the most concentrated known extraterrestrial occurrence of carbonate on kilometre-wide scales in the Solar System. The carbonates are mixed with a dark component and small amounts of phyllosilicates, as well as ammonium carbonate or ammonium chloride. Some of these compounds have also been detected in the plume of Saturn's sixth-largest moon Enceladus⁵. The compounds are endogenous and we propose that they are the solid residue of crystallization of brines and entrained altered solids that reached the surface from below. The heat source may have been transient (triggered by impact heating). Alternatively, internal temperatures may be above the eutectic temperature of subsurface brines, in which case fluids may exist at depth on Ceres today.

Occator is a prominent, geologically young crater^{1,6} (diameter ~90 km, 19.4° N, 239.0° E), and its brightest region (Fig. 1) is at the centre of the crater, close to a 10-km-wide central pit⁶. Several other smaller and less bright areas are present to the east on the crater floor. Occator has been observed in five spectral images (Fig. 1) from a distance of 1,400 km with a spatial resolution of about 380 m per pixel by the Visible and InfraRed Mapping Spectrometer (VIR)⁷ onboard the Dawn spacecraft. The analysis of spectra photometrically corrected to standard geometry indicates that the reflectance of most of Occator's dark floor (0.03 at 0.55 μ m and at 2 μ m) is similar to that of regions around the crater. The highest reflectance is measured in the central pit with a value of 0.26 at 0.55 μ m and 0.28 at 2 μ m. A transition zone with decreasing level of reflectance from the bright areas to the surroundings can be recognized, associated with a variation in composition (Fig. 2). The brightest areas show clear spectral differences from the typical crater floor (Fig. 2). The overall band area between 2.6 μ m and 3.7 μ m increases going from the crater floor to the brightest pixels, and most absorptions become deeper and better defined (Fig. 2a). The 2.7- μ m absorption of magnesium (Mg) phyllosilicates typical of Ceres' surface^{2,3} shifts from 2.72 μ m to 2.76 μ m (Fig. 2b). The 3.07- μ m absorption, attributed to ammonia-bearing species^{2–4}, is clearly present on the

crater floor but becomes less evident in brighter terrains and is absent in the brightest pixels (Fig. 2a and Fig. 3). Absorptions at 2.20–2.22 μ m and around 2.86 μ m appear only in the centre of the brightest areas (Figs 2b and c, and 3). Notably, absorptions near 3.4 μ m and 3.9 μ m increase markedly in the brightest areas and are consistent with enrichments in carbonates (Figs 2d and e, and 3), which have been previously identified as a common component of the surface of Ceres^{2,8}.

The bright material is clearly different from the rest of the crater floor and from the materials that typify most of the surface of Ceres^{2,4}. The dark floor material can be modelled with the same components used to describe the overall surface of Ceres² (Extended Data Fig. 1). In contrast, in high-albedo areas the relative proportions of the components changes, with a clear increase in carbonate content with respect to the dark material. Furthermore, additional phases appear. The shift of the metal hydroxide absorption from 2.72 μ m to 2.76 μ m (Fig. 2b) indicates a lower abundance of magnesium (Mg)-phyllosilicates and, possibly the occurrence of an aluminium (Al)-phyllosilicate such as a smectite, kaolinite or illite^{9,10}. Hexahydrite (MgSO₄·6H₂O) was initially suggested as a major component (30 vol%) of the bright areas on the basis of spectral variations at visible wavelengths¹. However, short-wavelength near-infrared data for such a case would also exhibit strong H₂O bands (Extended Data Fig. 2), which are now not observed in the Occator central pit spectra. Spectral modelling of the VIR data indicates that hexahydrite in amounts exceeding 2–3 vol% is inconsistent with these new VIR observations. Mixtures with water ice (Extended Data Fig. 3) indicate an upper limit of few volume per cent of water ice.

Strong carbonate bands, typical of anhydrous carbonates¹¹, dominate the spectra of Occator bright material near 3.4 μ m and 3.9 μ m (Fig. 2). Hydrous carbonates have weak-to-absent 3.4- μ m and 3.9- μ m features but strong H₂O bands¹², which are not found here. The carbonate absorption centre positions of Occator spectra show that band centres agree with those of natrite (Na₂CO₃) (Extended Data Fig. 4) in contrast to Ceres' average carbonate band centre, which is more consistent with dolomite CaMg(CO₃)₂. Moreover, the spectra of Occator bright material shows an absorption that extends to shorter wavelengths in the 3.2–3.5- μ m range with respect to that of Mg, Ca, Fe and Na carbonates, indicating the possible presence of other mineralogical phases. Many overlapping overtones and combinations of fundamental modes of CO₃^{2–} or X-H (for example, CH and OH) vibrations occur in the 3.2–3.5- μ m spectral range. Carbonaceous chondrites show organic-related absorptions in the 3.3–3.5- μ m range, but these absorptions do not extend to the shorter wavelengths observed in Occator bright material (Extended Data Fig. 5). However, uncertainties

¹Istituto di Astrofisica e Planetologia Spaziali, Istituto Nazionale di Astrofisica (INAF), Via del Fosso del Cavaliere 100, 00133 Rome, Italy. ²Earth Planetary and Space Sciences, University of California, Los Angeles, California, USA. ³Institut de Recherche d'Astrophysique et Planétologie, Université de Toulouse, Centre National de la Recherche Scientifique (CNRS), Université Paul Sabatier (UPS), Toulouse, France. ⁴Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, Tennessee 37996-1410, USA. ⁵Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA. ⁶Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA. ⁷Solar System Exploration Research Virtual Institute, Southwest Research Institute (SRI), 1050 Walnut Street, Boulder, Colorado 80302, USA. ⁸NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ⁹Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, Rhode Island 02912, USA. ¹⁰Institute of Planetary Research, German Aerospace Center (DLR), Rutherfordstrasse 2, 12489 Berlin, Germany. ¹¹Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, Texas 77058, USA. ¹²Agenzia Spaziale Italiana, Via del Politecnico, 00133 Roma, Rome, Italy.

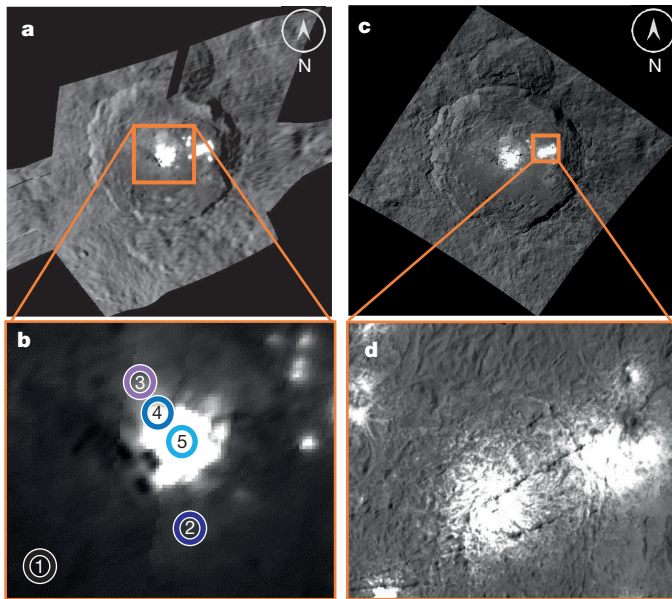


Figure 1 | Occator crater. **a**, Mosaic of Occator crater obtained with VIR data at 2 μm . The continuous colour from white to dark grey corresponds to reflectance from high (0.28) to low (0.03). **b**, Enlargement of the brightest central area. The coloured circled numbers correspond to the spectra in Fig. 2. **c**, Context image obtained with the Dawn Framing Camera during the High Altitude Mapping Orbits. **d**, Dawn Framing Camera image of the secondary bright areas acquired during the 375-km Low Altitude Mapping Orbits.

in the instrument calibration in the 3.2–3.4- μm spectral region call for caution when interpreting band shapes, and this part of the spectra will not be emphasized in this analysis.

The absorption at around 2.2 μm can be due to a narrow hydroxide (OH) stretching and bending combination vibration in Al-phylosilicates¹⁰ or to ammoniated minerals that also have bands at around 2.21 μm (refs 13–15), such as NH_4Cl and NH_4HCO_3 (Extended Data Fig. 6). The brightest regions in Occator (Fig. 3) are those where the 2.21- μm and the 3.9- μm bands are stronger and the 3.07- μm band is weaker, as shown

by the spectra and correlation trends (Figs 2 and 3d and e). Where the ammonium phyllosilicate band is weak or absent, another ammonium-bearing mineral (2.21- μm band) may be present, and the carbonate band at 3.9 μm is strongest.

Using these phase identifications, we modelled the Occator brightest material with a mixture of sodium carbonate (Na_2CO_3), low-albedo material, illite or montmorillonite, and a small quantity of ammonium-bearing species, including NH_4Cl or NH_4HCO_3 (Fig. 4). A common result of these mixture model fits is that carbonates are the most abundant species, (roughly 45–80 vol%, depending on the end-member spectra; see Extended Data Tables 1 and 2, and Methods). A small quantity of ammonium chlorides or bicarbonates and Al-phylosilicate (a few volume per cent), is also needed in these models to account for the band at 2.21 μm and the overall spectral shape (Extended Data Fig. 7).

Delivery of exogenous materials cannot account for the Occator bright areas, because the spectra are unlike other asteroids and comets. The morphologies of the bright areas (nearly circular shapes and correlation with fractures) argue against a direct impact origin, although the central mound of Occator and the presence along fractures (Fig. 1d) suggests that their emplacement may be related to the impact that formed Occator. However, it is unclear whether the bright material in Occator was formed from local aqueous processing triggered by the impact or if it represents an exposure of deeper material that found its way to the surface via fractures generated by that impact, or some combination of those two processes. In either case, these materials were derived from the interior of Ceres.

The central bright area in Occator is especially rich in carbonate, and it appears to represent the most concentrated known occurrence of km-scale carbonates beyond the Earth. Carbonates also occur in carbonaceous chondrites but in amounts of only a few volume per cent. Calcite, aragonite and dolomite together constitute <3 vol% of CM¹⁶ and CI¹⁷ chondrites. All these carbonates formed by aqueous alteration on their parent bodies within several Ma after formation of the solar system¹⁸. Despite the complexity of the carbonate mineralogy in CM and CI chondrites, no natrite has been reported.

On Ceres, the occurrence of ammonium salts, as well as the unexpectedly high proportion of carbonates in Occator bright materials, points to a formation mechanism that is distinct from that which produces meteoritic carbonates. On Earth natrite is a magmatic

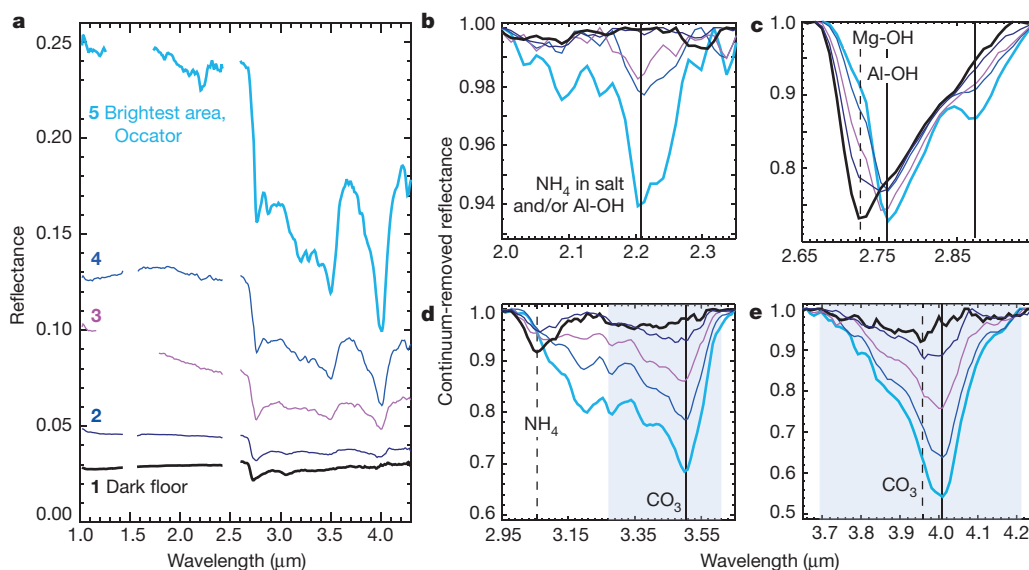


Figure 2 | Spectra of bright and dark areas in Occator. **a**, Reflectance spectra from the circled regions in Fig. 1a. Gaps correspond to removed instrumental artefacts or saturated channels. **b–e**, Continuum-removed spectra from selected wavelength regions. A 2.2- μm absorption is visible in the spectra of the brightest pixels. The 2.7- μm absorption shifts longward and an absorption at 2.87 μm appears, going from the crater floor to the

brightest pixels. The 3.05- μm absorption weakens, while the 3.4- μm and 3.9- μm absorptions strengthen in the brightest areas. The shaded region indicates the band positions and widths for anhydrous carbonates. Dashed lines indicate absorption positions in dark materials and solid lines indicate absorption positions in bright materials.

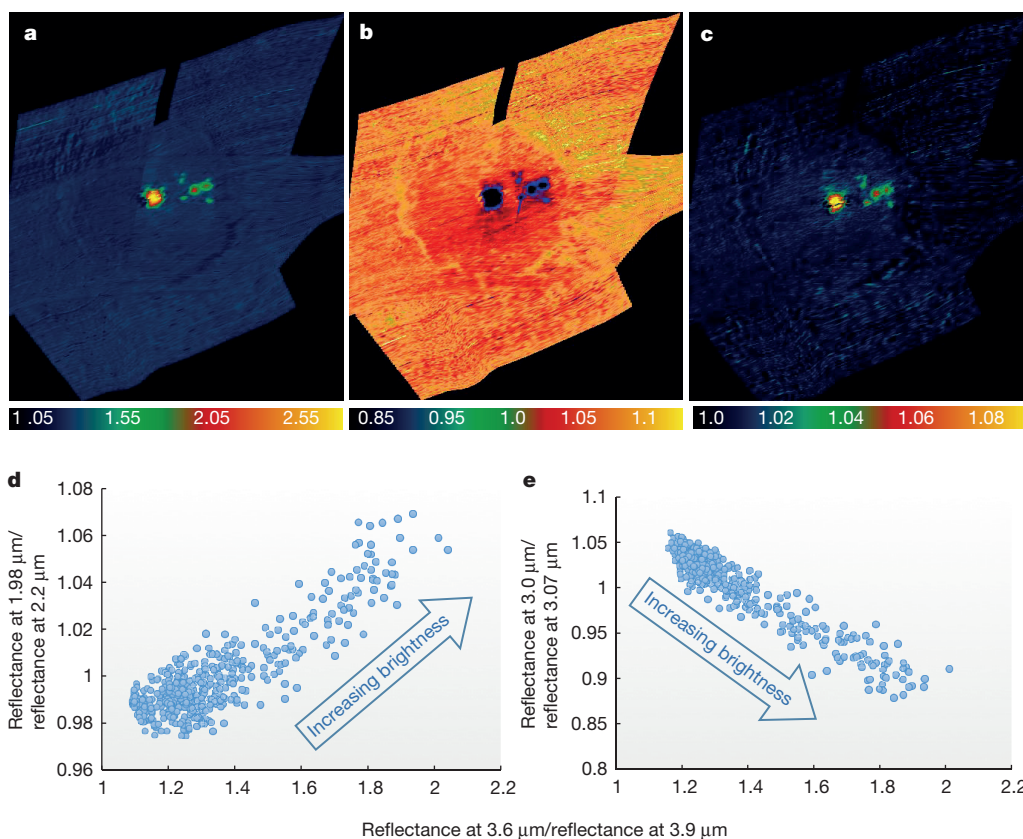


Figure 3 | Spatial distribution and scatter plots of different absorption intensities. a–c, The ratio between the reflectance at 3.6 μm and that at 3.9 μm (a) provides a proxy of the 3.9- μm band depth, the ratio between the reflectance at 3.0 μm and that at 3.07 μm (b) provides a proxy for the 3.07- μm band depth, and the ratio between the reflectance at 1.98 μm and that at 2.2 μm (c) provides a proxy of the 2.2- μm band depth. The colour

scale corresponds to the band depth (from blue at the minimum to yellow at the maximum). d, e, Scatter plots of band ratios show that the 3.9- μm band strength is correlated with the 2.21- μm band and anti-correlated with the 3.07- μm band depth. The 2.2- μm band is also anti-correlated with the 3.07- μm band.

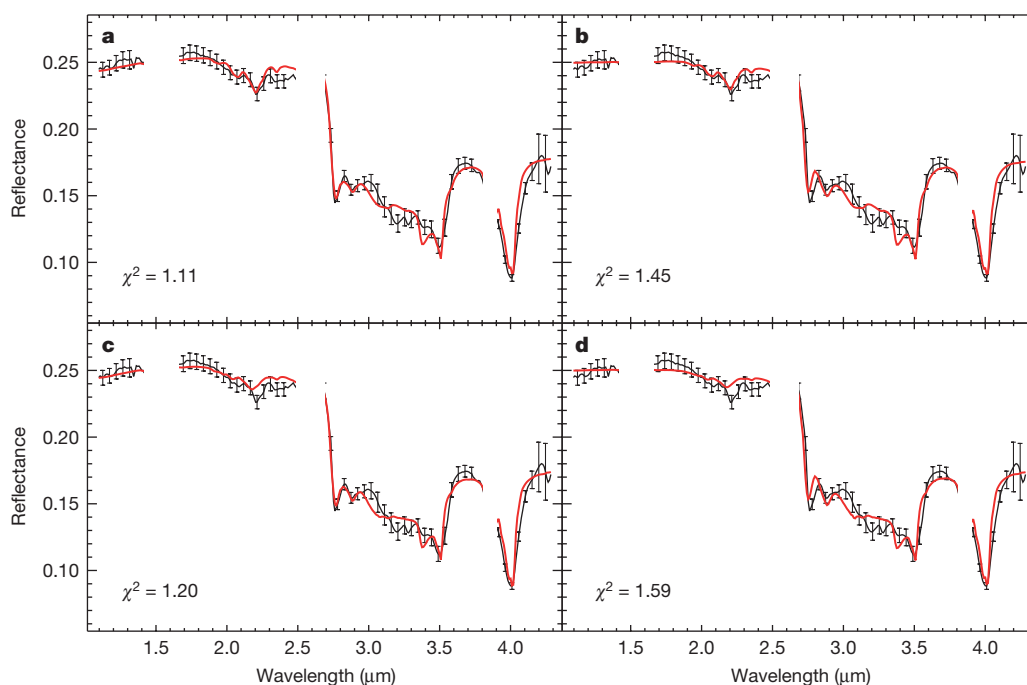


Figure 4 | Spectral fits of the Occator bright material spectrum. Results of the spectral fitting model (red line) using: a, dark material, natrite, illite and ammonium chloride; b, dark material, natrite, illite and ammonium bicarbonate; c, dark material, natrite, montmorillonite and ammonium chloride; and d, dark material, natrite, montmorillonite and ammonium

bicarbonate. Values of the computed χ^2 for each mixture are given in the plots. Error bars for the Occator spectrum (black) are calculated taking into account a mean absolute deviation of the calibration uncertainties along the 256 samples. The end-members are described in Extended Data Table 1 and the retrieved abundances in Extended Data Table 2.

mineral in certain carbonatites, but this origin is improbable on Ceres. Alternatively, natrite is found in terrestrial alkaline hydrothermal and evaporitic environments, and it has also been detected in Enceladus' plumes⁵, along with NaHCO_3 and NaCl . For the case of CI-chondritic type bodies, models for ammonia-free water-chondrite interaction^{19,20} suggest that Na-carbonate may be formed, but that the abundance of the latter is dominated by that of NaCl . In this respect, we note that NaCl is featureless in the spectral range observed by VIR, but that if present in the bright deposits of Occator, its abundance is small compared to that of carbonate. For this reason, we consider the role of ammonium that is known to compensate the acidifying action of dissolved CO_2 , keeping pH high enough for the stability of carbonate or bi-carbonate ions in the fluids. Indeed, if sodium carbonate dominates sodium chloride in the deposits on Ceres, as suggested here, a chemical pathway similar to that used to produce industrial Na_2CO_3 on Earth (the Solvay process) may be at work, where dissolved NaCl , CO_2 , and ammonia are used to precipitate sodium bicarbonate, producing ammonium chloride and/or bicarbonate as possible reaction products. Support for this sort of chemical scenario on Ceres is provided by the ubiquitous presence of ammonium-bearing minerals on Ceres⁶, the presence of minor ammonium salts in Occator, and the fact that abundant carbon is expected on bodies that resemble carbonaceous chondrites in bulk composition.

Occator is a fresh crater, possibly 100 million years or younger in age^{1,6}, so the bright spots should be of similar or younger age. However, the source of modern fluids on Ceres remains an open question. When dissolved in water, NaHCO_3 (the precursor of Na_2CO_3) has a eutectic point of $\sim 267\text{ K}$, while NH_4Cl and NH_4HCO_3 have eutectics at $\sim 251\text{ K}$ and $\sim 256\text{ K}$, respectively (see Methods). These eutectic temperatures lie above the maximum temperature predicted for the outermost 100 km or so of Ceres' crust, in which case these materials should remain solid²¹. The occurrence of carbonates and other salt assemblages in Occator may thus point to internal temperatures that are warmer than predicted by the models (for example, because of the low thermal conductivities of salts). Alternatively, an external heat source may have been involved, such as impact-induced heating. Indeed, estimates suggest that the impact that formed Occator could have increased the surface temperature up to that of the melting point of water ice²², hence encompassing the eutectics of the aforementioned species.

The morphology of the bright areas indicates an association with fracture systems that may have facilitated brine upwelling (Fig. 1d). These fractures were either created by impacts, or they may be associated with subsequent internal movements. In either case, upon ascent and exposure the solute-bearing fluid containing entrained altered solids froze, causing precipitation and concentration of carbonates and salts. The detection of abundant sodium carbonate, albeit in localized regions at the surface of Ceres, provides constraints on Ceres' chemical evolution and indicates that aqueous alkaline solutions could even persist in Ceres' subsurface to the present day. In addition, these regions bear similarities with Enceladus, where ammonia, NaCl , NaHCO_3 and Na_2CO_3 have been detected in plumes²³. These observations point to Ceres as an object that has experienced aqueous processes in the recent geological past involving materials similar to those predicted or observed on icy satellites^{5,20,23,24}, confirming the link of Ceres with the bodies of the outer Solar System².

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 February; accepted 25 April 2016.

Published online 29 June 2016.

1. Nathues, A. *et al.* Sublimation in bright spots on (1) Ceres. *Nature* **528**, 237–240 (2015).
2. De Sanctis, M. C. *et al.* Ammoniated phyllosilicates with a likely outer Solar System origin on (1) Ceres. *Nature* **528**, 241–244 (2015).

3. King, T., Clark, R., Calvin, W., Sherman, D. & Brown, R. Evidence for ammonium-bearing minerals on Ceres. *Science* **255**, 1551–1553 (1992).
4. Ammannito, E. *et al.* Distribution of phyllosilicates on Ceres. *Science* (in the press).
5. Postberg, F., Schmidt, J., Hillier, J., Kempf, S. & Srama, R. A salt-water reservoir as the source of a compositionally stratified plume on Enceladus. *Nature* **474**, 620–622 (2011).
6. Buczkowski, D. *et al.* The geomorphology of Ceres. *Science* (in the press).
7. De Sanctis, M. C. *et al.* The VIR spectrometer. *Space Sci. Rev.* **163**, 329–369 (2011).
8. Rivkin, A. S., Volquardsen, E. L. & Clark, B. E. The surface composition of Ceres: discovery of carbonates and iron-rich clays. *Icarus* **185**, 563–567 (2006).
9. Beran, A. in *Micas: Crystal Chemistry and Metamorphic Petrology* (eds Mottana, A., Sassi, F. P., Thompson, J. B. & Guggenheim, S.). 351–370 (Mineralogical Society of America, 2002).
10. Bishop, J. L. *et al.* Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. *Clay Miner.* **43**, 35–54 (2008).
11. Hunt, G. R. & Salisbury, J. W. Visible and near infrared spectra of minerals and rocks. II. Carbonates. *Mod. Geol.* **2**, 23–30 (1971).
12. Harner, P. L. & Gilmore, M. S. Visible-near infrared spectra of hydrous carbonates, with implications for the detection of carbonates in hyperspectral data of Mars. *Icarus* **250**, 204–214 (2015).
13. Moore, M. H. *et al.* Ammonia water ice laboratory studies relevant to outer Solar System surfaces. *Icarus* **190**, 260–273 (2007).
14. Moore, M. H., Hudson, R. L. & Ferrante, R. F. Radiation products in processed ices relevant to Edgeworth-Kuiper-belt objects. *Earth Moon Planets* **92**, 291–306 (2003).
15. Berg, B. L. *et al.* Reflectance spectroscopy (0.35–8 μm) of ammonium-bearing minerals and qualitative comparison to Ceres-like asteroid. *Icarus* **265**, 218–237 (2016).
16. de Leuw, S., Rubin, A. E. & Wasson, J. T. Carbonates in CM chondrites: complex formational histories and comparison to carbonates in CI chondrites. *Meteorit. Planet. Sci.* **45**, 513–530 (2010).
17. Johnson, C. A. & Prinz, M. Carbonate compositions in CM and CI chondrites and implications for aqueous alteration. *Geochim. Cosmochim. Acta* **57**, 2843–2852 (1993).
18. Brearley, A. J. in *Meteorites and the Early Solar System II* 587–624 (Univ. Arizona Press, 2006).
19. Zolotov, M. Yu. Aqueous fluid composition in CI chondritic materials: chemical equilibrium assessments in closed systems. *Icarus* **220**, 713–729 (2012).
20. Zolotov, M. Y. An oceanic composition on early and today's Enceladus. *Geophys. Res. Lett.* **34**, L23203 (2007).
21. Castillo-Rogez, J. C. & McCord, T. B. Ceres' evolution and present state constrained by shape data. *Icarus* **205**, 443–459 (2010).
22. Bowling, T. J. *et al.* Impact induced heating of Occator crater on asteroid 1 Ceres. *47th Lunar Planetary Sci. Conf. abstr.* 2268 (2016).
23. Waite, J. H. *et al.* Liquid water on Enceladus from observations of ammonia and ^{40}Ar in the plume. *Nature* **460**, 487–490 (2009).
24. Postberg, F. *et al.* Sodium salts in E ring ice grains from an ocean below the surface of Enceladus. *Nature* **459**, 1098–1101 (2009).

Acknowledgements We thank the following institutions and agencies which supported this work: the Italian Space Agency, the National Aeronautics and Space Administration (NASA, USA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR, Germany). The VIR was funded and coordinated by the Italian Space Agency and built by SELEX ES, with the scientific leadership of the Institute for Space Astrophysics and Planetology and the Italian National Institute for Astrophysics, and is operated by the Institute for Space Astrophysics and Planetology, Italy. A portion of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, USA, under contract to NASA. We also thank the Dawn Mission Operations team and the Framing Camera team.

Author Contributions M.C.D.S., A.R., E.A. and F.G.C. performed data analysis and calibration. M.C. provided optical constants from reflectance spectra. M.C.D.S., C.M.P. and B.L.E. contributed to the spectral interpretation of the data. All authors contributed to the discussion of the results and to writing the paper.

Author Information The VIR calibrated data will be made available through the PDS Small Bodies Node website (<http://sbn.pds.nasa.gov/>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.D.S. (mariacristina.desanctis@iaps.inaf.it).

Reviewer Information *Nature* thanks V. Reddy, A. S. Rivkin and M. M. Zolotov for their contribution to the peer review of this work.

METHODS

Composition analysis. The spectra analysed here have been corrected for standard geometry (incidence angle 30° , emission angle 0° , phase angle 30°) to derive the values of reflectance. In particular, the central Occator bright spectrum reported in the main text is derived as an average of the four brightest spectra acquired during acquisition number 498468846 with the following viewing geometry: phase angle 45° , incidence angle 18° , emission angle 41° . The resulting spectrum has also been corrected for thermal emission.

We model Ceres' average spectrum as an intimate mixture of different end-members, by means of Hapke theory²⁵, which characterizes light scattering in particulate media. The optical constants are derived from reflectance spectra as in ref. 26. Abundances and grain size of the components are free parameters, and a best fit is obtained by means of a least-squares optimization algorithm^{27–29}. The model accounts for the viewing geometry (incidence, emission and phase angle) which is calculated according to the shape model, the spacecraft attitude, and the latitude and longitude of a given pixel on the surface. The model also accounts for the single particle phase function, which is a free parameter in the fitting procedure.

We take advantage of the modelling performed for the overall surface of Ceres⁶ as a starting point for the bright spot composition. A first fit is made by adding water ice to the end-members listed in Extended Data Table 1 (magnetite, antigorite, NH_4 -montmorillonite and dolomite). Although the model can roughly account for the increase in the depth of the complex $3\text{-}\mu\text{m}$ band, this fit is unsatisfactory. Absorption bands of carbonates ($3.5\text{-}\mu\text{m}$, $4.0\text{-}\mu\text{m}$) are poorly fitted; other minor absorption bands in the measured spectrum ($2.2\text{-}\mu\text{m}$, $2.9\text{-}\mu\text{m}$) are absent in the model; conversely, some bands required by the model, like the $2.0\text{-}\mu\text{m}$ absorption band of water ice, have not been observed.

As discussed in the main text, the best fit is obtained by changing the kind of carbonate and phyllosilicate with respect to the average composition of the surface² and adding ammonium chloride, or ammonium bicarbonate (Extended Data Table 1 and 2). The components used to produce the best fit (Fig. 4) and the references for the spectra used for the analysis are listed in Extended Data Table 1. The model with water ice^{30–33} does not improve the fits. However, a small amount (1 vol%) is still compatible (Extended Data Table 2). The parameters obtained are listed in Extended Data Table 2.

In the case of dark materials, the albedo is weakly linked to grain size, so grain size cannot be determined unequivocally. However, it can be constrained to be within the range $20\text{--}100\text{-}\mu\text{m}$, which is reasonable for surface regolith. For each combination of end-members, the best fit is repeated for the two grain size limits of the dark material. From the distribution of the results (Extended Data Table 2), one can evaluate the mean value and the uncertainties on the retrieved parameters.

Formation of Occator's bright area material. The abundant carbonate in Occator's bright area requires (i) substantial amounts of carbon in solution, perhaps derived from the dissolution of organics, which are found in carbonaceous chondrites, and potentially augmented by CO/CO_2 -bearing ices that may have accompanied the nitrogen-bearing material from the outer Solar System²; (ii) substantial amounts of fluid with dilute carbonate and bicarbonate ions, and/or (iii) a concentration of carbonate at or near the surface by migration and evaporation of fluids near the

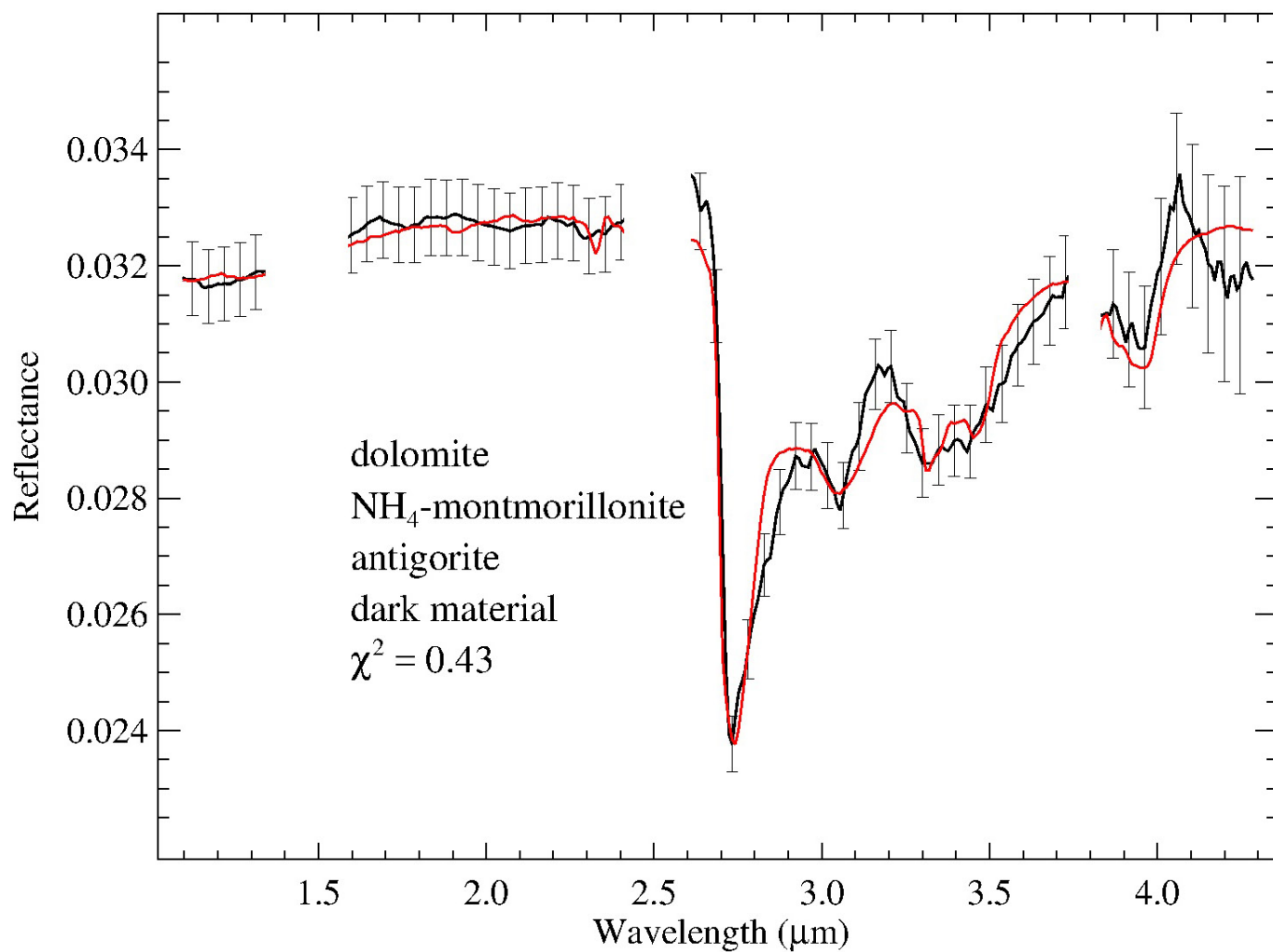
surface of Ceres, similar to evaporates or 'caliche' in terrestrial soils. Ammonium may be present in the form of ammoniated salts from freezing of the early ocean in which the observed surface mineralogy formed, or it may be derived from brines interacting with the NH_4 present in typical Ceres surface materials².

Ammonium makes the solution basic, converting dissolved carbonate anions into bicarbonate anions, which then readily associate with ammonium cations. However, in the presence of dissolved NaCl , the ammonium preferentially combines with chlorine while sodium combines with bicarbonate. This assemblage can be exposed on the surface of Ceres, through fractures or cratering events, following emplacement and sublimation of the ice.

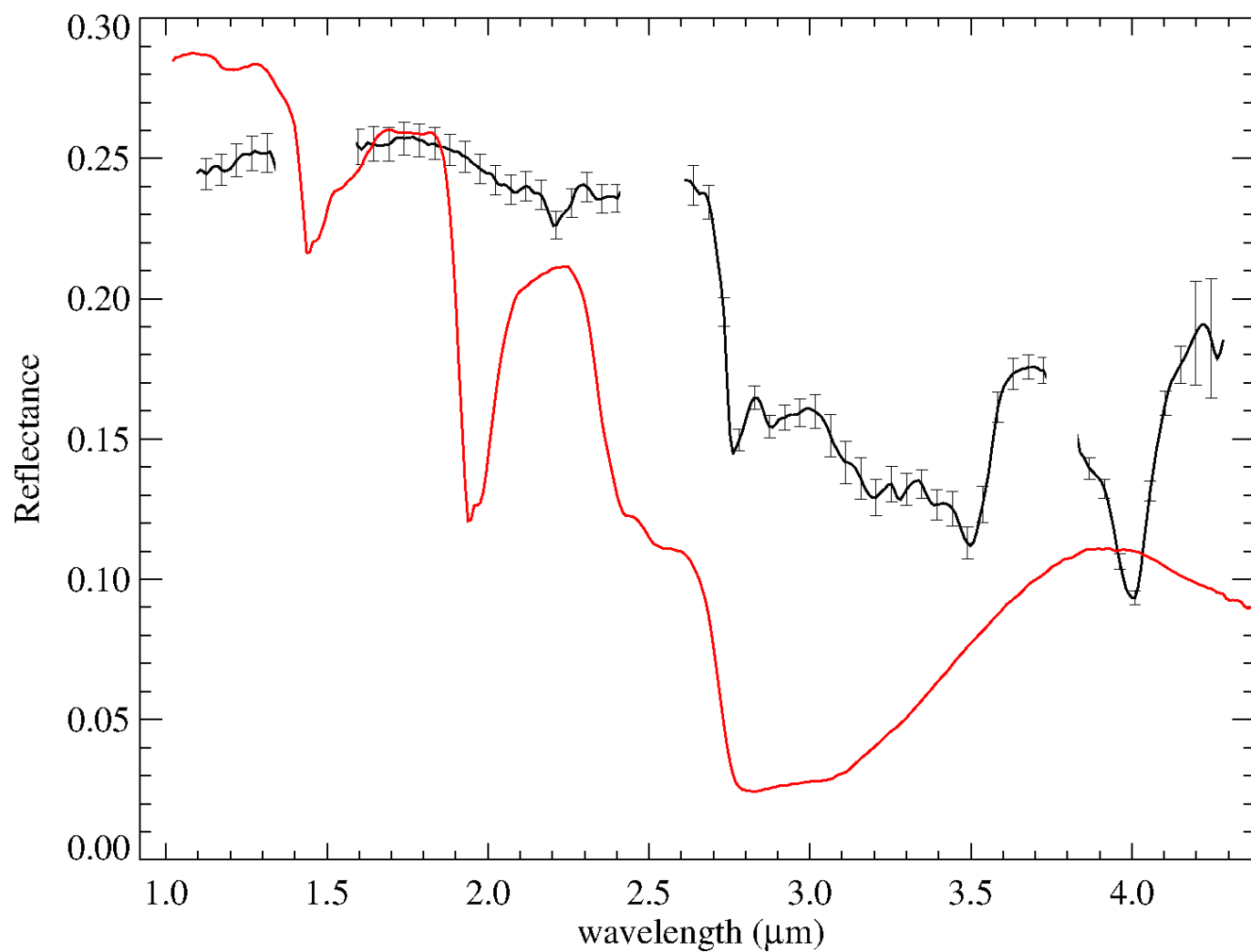
The association of sodium carbonate and ammonium chloride (and/or ammonium carbonate) described in the Occator bright spots is thus consistent with the freezing or evaporation products of a salt solution bearing carbon species and ammonium ions. This process is similar to that employed for the synthesis of industrial sodium carbonate on Earth³⁴, except that the saturation of the solution in CO_2 does not require the high-temperature breakdown of calcite or limestone if alternative sources of carbon are already present in solution.

Code availability. We have opted not to make the code available for fitting the data because it is a numerical code developed specifically for this purpose, but the code is described in refs 27, 28 and 29.

25. Hapke, B. *Theory of Reflectance and Emissance Spectroscopy* (Cambridge Univ. Press, 2012).
26. Carli, C., Ciarniello, M., Capaccioni, F., Serventi, G. & Sgavetti, M. Spectral variability of plagioclase-mafic mixtures (2): investigation of the optical constant and retrieved mineral abundance dependence on particle size distribution. *Icarus* **235**, 207–219 (2014).
27. Raponi, A. Spectrophotometric analysis of cometary nuclei from in situ observations. PhD thesis, Univ. degli studi di Roma Tor Vergata; preprint at <http://arxiv.org/abs/1503.08172> (2015).
28. De Sanctis, M. C. *et al.* The diurnal cycle of water ice on comet 67P/Churyumov-Gerasimenko. *Nature* **525**, 500–503 (2015).
29. Filacchione, G. *et al.* Exposed water ice on the nucleus of comet 67P/Churyumov-Gerasimenko. *Nature* **529**, 368–372 (2016).
30. Warren, S. G. Optical constants of ice from the ultraviolet to the microwave. *Appl. Opt.* **23**, 1206–1225 (1984).
31. Mastrapa, R. M. *et al.* Optical constants of amorphous and crystalline H_2O -ice in the near infrared from 1.1 to $2.6\text{-}\mu\text{m}$. *Icarus* **197**, 307–320 (2008).
32. Mastrapa, R. M. *et al.* Optical constants of amorphous and crystalline H_2O -ice: $2.5\text{--}22\text{-}\mu\text{m}$ ($4000\text{--}455\text{-cm}^{-1}$) optical constants of H_2O -ice. *Astrophys. J.* **701**, 1347–1356 (2009).
33. Clark, R. N. *et al.* The surface composition of Iapetus: mapping results from Cassini VIMS. *Icarus* **218**, 831–860 (2012).
34. Kiefer, D. M. Soda ash, Solvay style. *Today's Chemist at Work* **11** (2), 87–88, 90 (2002).
35. Clark, R. N. *et al.* Splib06a. *US Geol. Surv. Data* **231**, <http://speclab.cr.usgs.gov/spectral.lib06/> (USGS Digital Spectral Library, 2007).
36. Takir, D. *et al.* Nature and degree of aqueous alteration in CM and CI carbonaceous chondrites. *Meteorit. Planet. Sci.* **48**, 1618–1637 (2013).
37. Milliken, R. E. & Mustard, J. F. Quantifying absolute water content of minerals using near infrared reflectance spectroscopy. *J. Geophys. Res.* **110**, E12001 (2005).

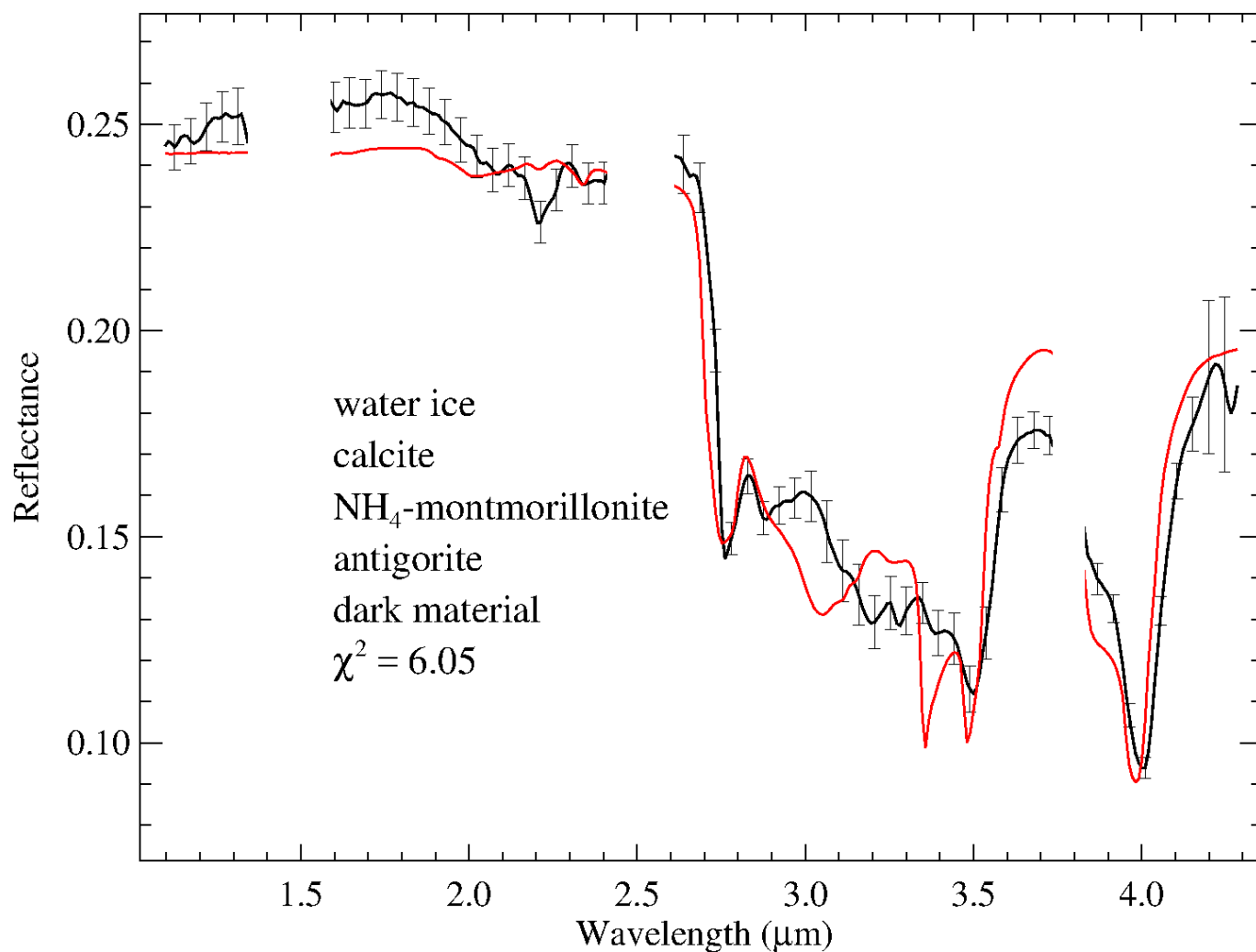


Extended Data Figure 1 | Occator floor spectrum. Spectral fit (red) of the reflectance spectrum of the floor of the Occator crater (black) using the same end-members as discussed for the average surface². The computed χ^2 for the mixture is given in the plot. Error bars for the Occator floor spectrum are calculated taking into account a mean absolute deviation of the calibration uncertainties along the 256 samples.



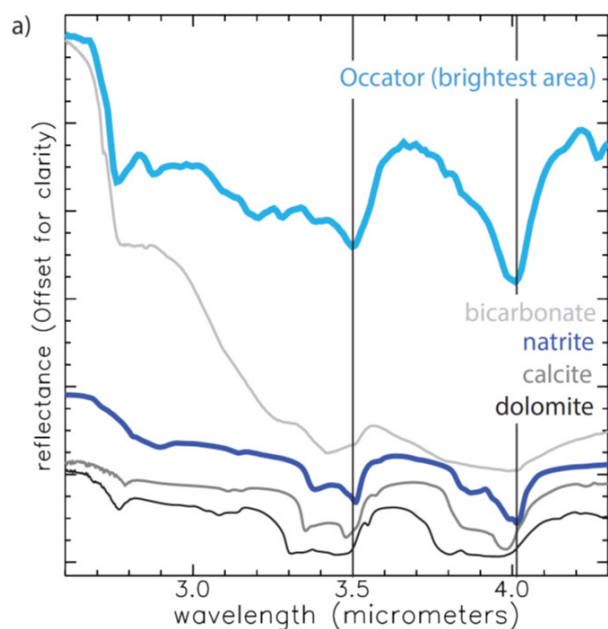
Extended Data Figure 2 | Spectral fit with hexahydrite. Modelled spectrum (red) of a mixture of hexahydrite ($\text{MgSO}_4 \cdot 6\text{H}_2\text{O}$) (30 vol%) with the average for Ceres (70 vol%). Strong absorptions bands due to H_2O are visible at 1.4 μm , 1.95 μm , 2.45 μm and 3 μm in the modelled spectrum,

which are not observed in the Occator bright spectrum (black). Error bars for the Occator spectrum are calculated taking into account a mean absolute deviation of the calibration uncertainties along the 256 samples.

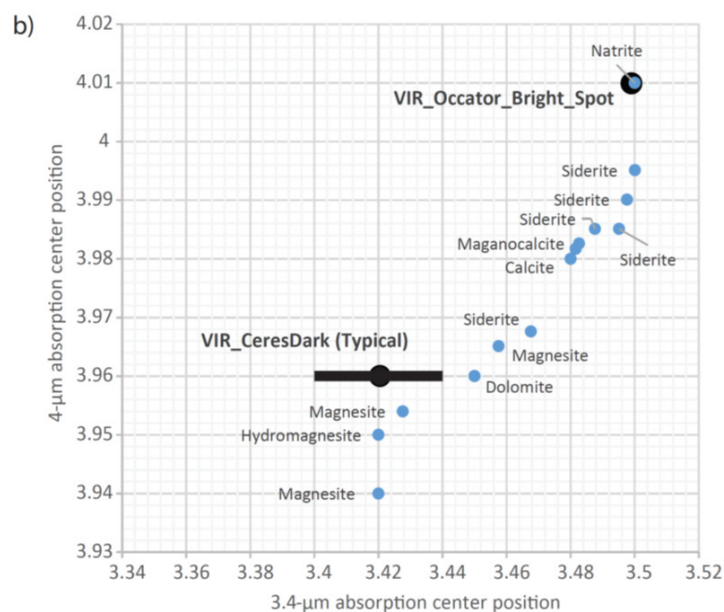


Extended Data Figure 3 | Spectral fit with water ice. Spectral fit (red) of the reflectance spectrum of the bright spot in Occator (black) using the same end-members discussed for the average surface² and adding water ice (Extended Data Table 1). Resulting parameters are reported in Extended Data Table 2. Several absorptions present are poorly fitted, and several

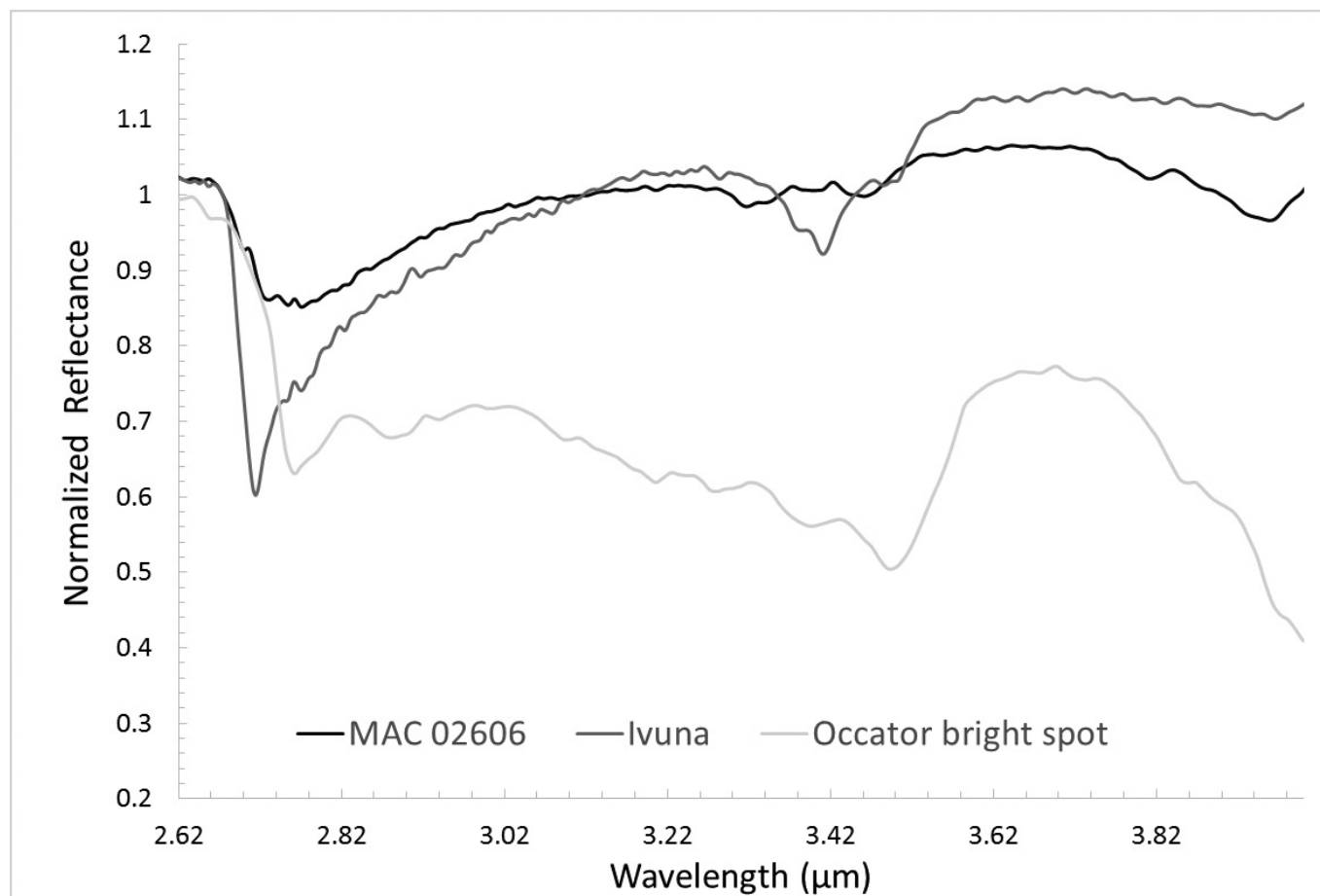
absorptions predicted by the best-fit model are absent. The computed χ^2 for the mixture is given in the plot. Error bars for the Occator spectrum are calculated taking into account a mean absolute deviation of the calibration uncertainties along the 256 samples.



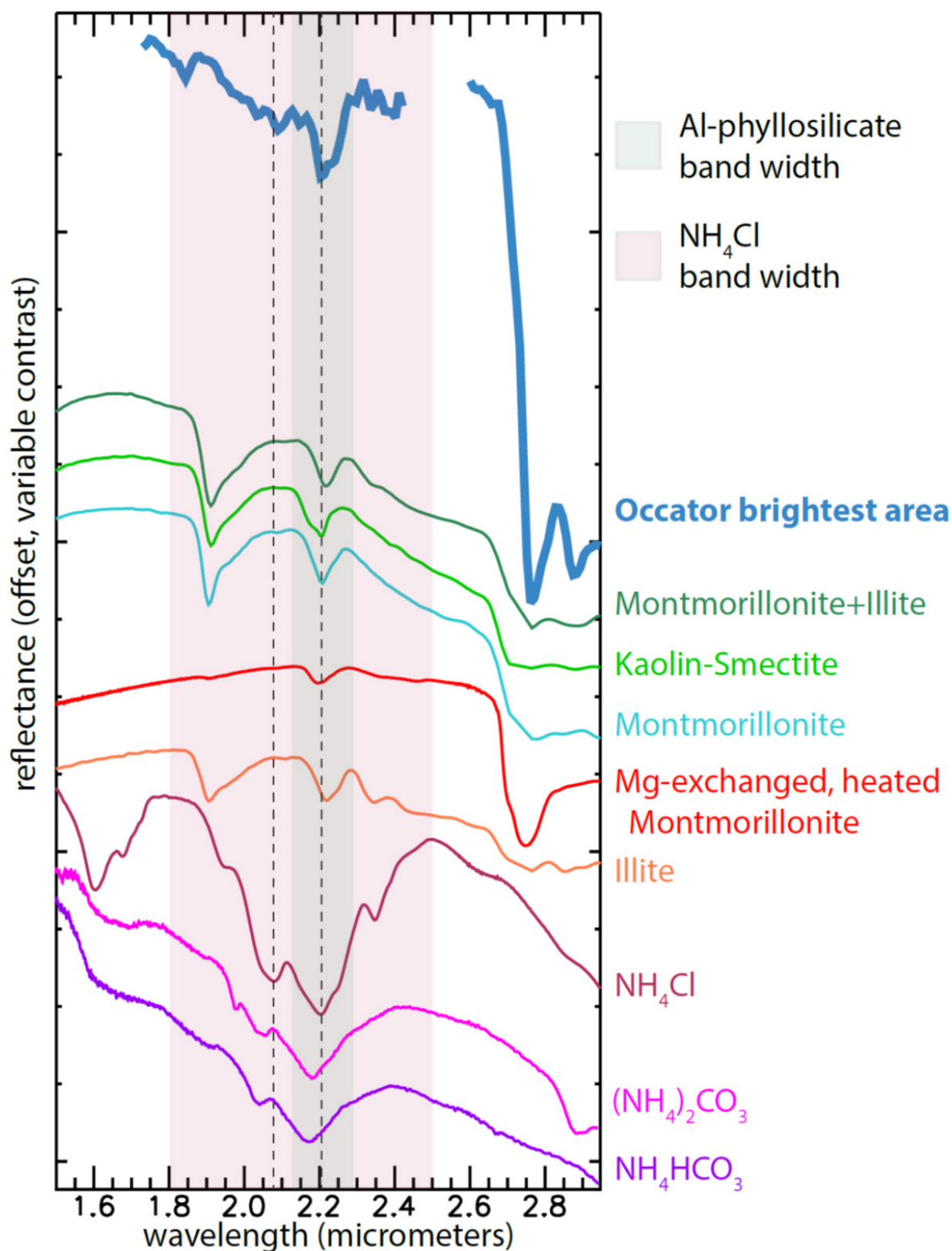
Extended Data Figure 4 | Comparison of Occator bright material spectrum with carbonates. **a**, Continuum-removed spectrum of the Occator bright material compared to natrite¹⁵, sodium bicarbonate (see <http://psf.uwinnipeg.ca>), calcite³⁵, and dolomite³⁵. **b**, A scatter plot of the longest-wavelength continuum-removed absorption band centres for different carbonates shows that Ceres data from Occator bright areas are similar to data from natrite and are distinct from data from other



parts of the planet, which plot near magnesite. The 3.9-μm absorption is strong in both the Occator bright areas and in dark floor material. The 3.4-μm absorption is strong in the bright areas but broader; its centre is challenging to define over most of the Ceres surface owing to the presence of other optically active phases. Spectral sampling of laboratory data varied, but in all cases was $<0.01 \mu\text{m}$. Error bars for the Occator spectrum are not reported in the plot.

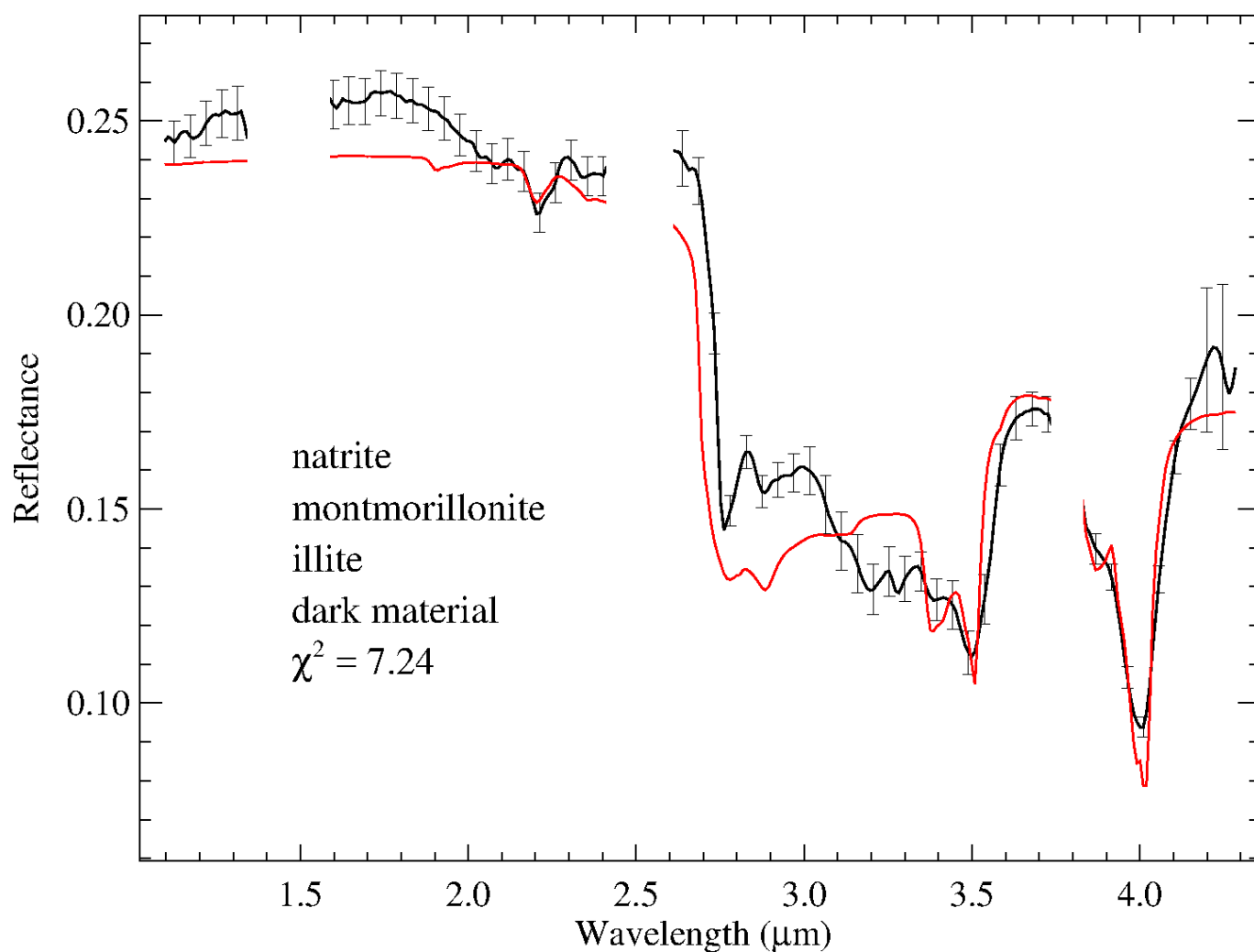


Extended Data Figure 5 | Comparison of Occator bright material spectrum with carbonaceous chondrites. Comparison of spectra from two carbonate- and organic-bearing carbonaceous chondrites (MAC 02606 (CM2) and Ivuna (CI))³⁶ and the Occator bright material. The spectra have been normalized to 1 at 2.62 μm. The Occator spectrum is not reported in the plot.



Extended Data Figure 6 | Spectra of candidate materials fitting the observed 2.20–2.22- μm absorption in Occator bright materials. The shaded grey area indicates the expected bandwidth of aluminium phyllosilicates³⁵ (see the PSF web site at <http://psf.uwinnipeg.ca> and the RELAB database at <http://www.planetary.brown.edu/rehab/>), and the

shaded red area indicates the expected bandwidth of NH_4Cl (ref. 15). The Occator spectrum is 20 \times contrast-enhanced. The magnesium-exchanged montmorillonite is heated to 300 $^\circ\text{C}$ (ref. 37). Spectra of ammoniated salts NH_4Cl , $(\text{NH}_4)_2\text{CO}_3$ and NH_4HCO_3 (ref. 15) are also plotted. Dotted lines correspond to ammonium absorptions near 2 μm .



Extended Data Figure 7 | Spectral fit without NH_4 salt. Spectral fit (red) of reflectance spectrum of the bright spot in Occator (black) using the same end-members as discussed for the natrite, montmorillonite, illite and dark material (Extended Data Table 1). Resulting parameters are reported

in Extended Data Table 2. The computed χ^2 for the mixture is given in the plot. Error bars for the Occator spectrum are calculated taking into account a mean absolute deviation of the calibration uncertainties along the 256 samples.

Extended Data Table 1 | End-members used in evaluating mixing model results

Mineral ¹	Sample ID	Spectrum ID
Murchison IOM	OG-CMA-002	BKR1OG002
magnetite	MG-EAC-002	LAMG02
natrite (1)	CB-EAC-034-C	LACB34C
natrite (2)	CB-EAC-079-A	BKR1CB079A
illite	IL-EAC-001	LAIL01
montmorillonite	EA-EAC-028-A	BKR1EA028A
NH ₄ Cl	CL-EAC-049-A	LACL49A
NH ₄ HCO ₃	CB-EAC-041-B	LACB41B
antigorite	AT-TXH-007	LAAT07
NH ₄ -montmorillonite	JB-JLB-189	397F189
dolomite	CB-EAC-003	LACB03A
hexahydrite	SF-EAC-057A	LASF57A
water ice		

¹Spectra have been selected from the RELAB database (<http://www.planetary.brown.edu/rehab/>). Natrite (1) and (2) are both used in the fitting procedure. IOM, insoluble organic matter. Optical constants for water ice are from refs 30–33.

Extended Data Table 2 | Combination of end-members used to produce the best fit

Endmembers	Cross section (%)	Grain size (μm)	Volume (%)	χ^2	Figure
water ice calcite NH ₄ -montmorillonite antigorite dark material	8 - 8 37 - 37 17 - 19 6 - 3 31 - 32	7 - 8 100 - 100 50 - 50 100 - 100 20 - 100	1 - 1 64 - 45 15 - 11 10 - 4 11 - 38	7.01 - 6.05	Fig. 3 extended data
natrite montmorillonite illite dark material	46 - 48 16 - 16 11 - 12 27 - 24	54 - 49 100 - 100 5 - 5 20 - 100	53 - 37 34 - 25 1 - 1 11 - 37	8.25 - 7.24	Fig. 7 extended data
natrite ammonium bicarbonate montmorillonite dark material	39 - 39 7 - 7 30 - 31 24 - 24	57 - 63 59 - 35 5 - 5 20 - 100	67 - 47 13 - 4 5 - 3 15 - 45	1.72 - 1.59	Fig. 4 main text
natrite ammonium bicarbonate illite dark material	36 - 35 7 - 6 39 - 41 19 - 18	48 - 51 74 - 47 8 - 8 20 - 100	59 - 43 17 - 7 11 - 8 13 - 43	1.37 - 1.20	Fig. 4 main text
natrite ammonium chloride montmorillonite dark material	38 - 39 5 - 6 32 - 31 24 - 24	83 - 84 26 - 15 5 - 5 20 - 100	80 - 56 3 - 1 4 - 3 12 - 40	1.63 - 1.45	Fig. 4 main text
natrite ammonium chloride illite dark material	34 - 34 4 - 5 43 - 43 19 - 18	80 - 76 36 - 21 7 - 7 20 - 100	76 - 54 4 - 2 9 - 7 11 - 37	1.46 - 1.11	Fig. 4 main text
water ice natrite ammonium chloride illite dark material	3 - 3 33 - 33 5 - 5 40 - 41 19 - 18	7 - 6 84 - 82 36 - 23 9 - 8 20 - 100	0.5 - 0.5 76 - 55 4 - 2 9 - 6 10 - 36	1.36 - 1.05	

The cross-section fraction and the grain size are free to vary in order to obtain the best fit. For each combination we obtain two different solutions related to the grain size of the dark material (20 μm and 100 μm), which is the only parameter fixed in the fitting procedure. The volume fraction is obtained according to the cross-section and the grain size of each end-member.

Controlling charge quantization with quantum fluctuations

S. Jezouin^{1*}, Z. Iftikhar^{1*}, A. Anthore¹, F. D. Parmentier¹, U. Gennser¹, A. Cavanna¹, A. Ouerghi¹, I. P. Levkivskyi², E. Idrisov³, E. V. Sukhorukov³, L. I. Glazman⁴ & F. Pierre¹

In 1909, Millikan showed that the charge of electrically isolated systems is quantized in units of the elementary electron charge e . Today, the persistence of charge quantization in small, weakly connected conductors allows for circuits in which single electrons are manipulated, with applications in, for example, metrology, detectors and thermometry^{1–5}. However, as the connection strength is increased, the discreteness of charge is progressively reduced by quantum fluctuations. Here we report the full quantum control and characterization of charge quantization. By using semiconductor-based tunable elemental conduction channels to connect a micrometre-scale metallic island to a circuit, we explore the complete evolution of charge quantization while scanning the entire range of connection strengths, from a very weak (tunnel) to a perfect (ballistic) contact. We observe, when approaching the ballistic limit, that charge quantization is destroyed by quantum fluctuations, and scales as the square root of the residual probability for an electron to be reflected across the quantum channel; this scaling also applies beyond the different regimes of connection strength currently accessible to theory^{6–8}. At increased temperatures, the thermal fluctuations result in an exponential suppression of charge quantization and in a universal square-root scaling, valid for all connection strengths, in agreement with expectations⁸. Besides being pertinent for the improvement of single-electron circuits and their applications, and for the metal–semiconductor hybrids relevant to topological quantum computing⁹, knowledge of the quantum laws of electricity will be essential for the quantum engineering of future nanoelectronic devices.

Some of the most fundamental theoretical predictions have so far eluded experimental confirmation. Charging effects are generally found to diminish as the conductances of the contacts are increased^{10–18}; however, although some measurements support the fundamental prediction^{6–8} that charge quantization vanishes in the presence of one ballistic channel^{10–12,17}, others conclude the opposite^{18–23}. Unsurprisingly, the scaling behaviour predicted for the reduction of charge quantization^{6–8} has also remained elusive, until now, despite several attempts^{16,17}.

A plausible explanation of the varying results regarding the charge quantization criteria is that, in the previously investigated devices, the quantum channels and the conductor were not completely distinct circuit elements. With a small island, in which the density of states is discrete, the non-local electronic wave functions merge the connected channels and the island into a complex quantum conductor, where Coulomb interactions may play a non-trivial role. As a result, charging effects can develop even if one of the conduction channels taken separately is perfectly ballistic. This phenomenon is called mesoscopic Coulomb blockade^{18,22,24}.

Investigating charge quantization at the most elemental single-channel level therefore requires tunable conduction channels linked

to a conductor with a negligible electronic level spacing. Although this can be realized by making the island larger, its size must remain small enough to preserve charge quantization. Indeed, thermal fluctuations average out charge quantization unless the charging energy associated with the addition of one electron in the island— $E_C = e^2/2C$, where the geometrical capacitance of the island C increases with size—is larger than the thermal energy $k_B T$, with k_B the Boltzmann constant and T the temperature^{1,2}.

We have solved these conflicting requirements with the hybrid metal–semiconductor single-electron transistor (SET) shown in Fig. 1a, implementing the schematic circuit of Fig. 1b: a central metallic island with a continuous density of states (coloured red in Fig. 1a, b) is connected to large electrodes (represented by white disks) through two Ga(Al)As quantum point contacts (QPC_{L,R}) that emulate single-channel quantum conductors over the entire range of coupling strengths.

The metallic island, which is made of a metallic AuGeNi alloy, has a negligible electronic level spacing $\delta \approx k_B \times 0.2 \mu\text{K}$, five orders of magnitude smaller than the base electronic temperature $T \approx 17 \text{ mK}$. It is galvanically connected, by thermal annealing, to a 105-nm-deep, Ga(Al)As, high-mobility two-dimensional electron gas (2DEG; darker grey areas delimited by bright lines in Fig. 1a). Achieving an almost perfectly transparent metal–2DEG electrical contact is crucial to reach the ballistic channel limit. Remarkably, the reflection probability of electrons at the interface is below 0.05%.

The QPCs are located in the 2DEG and tuned by field effect with the voltage applied to capacitively coupled metallic split gates (coloured green in Fig. 1a; the top-right split gates that are coloured yellow are negatively biased to remove the 2DEG underneath). Besides tuning, the precise characterization of each QPC, independently, is necessary for the quantitative exploration of charge quantization versus connection strength. However, in the SET configuration, the QPC conductances are interconnected and renormalized by Coulomb blockade. Moreover, only their series combination is accessible. To completely characterize QPC_{L,R}, we implemented with adjacent gates (coloured blue in Fig. 1a) the on-chip switches shown in Fig. 1b. The measured quantities $\tau_{L,R} \equiv G_{L,R}^{\text{QPC}} h/e^2$ (with h the Planck constant and $G_{L,R}^{\text{QPC}}$ the conductances of QPC_{L,R} when switches are closed (inset of Fig. 1c)) directly give the ‘intrinsic’ (not renormalized by Coulomb blockade) electron transmission probabilities of the constitutive quantum channels, which fully characterize the connection strength to the metallic island. As illustrated in Fig. 1c, $\tau_{L(R)} \leq 1$ corresponds to a single (spin-polarized, see below) channel of transmission probability $\tau_{L(R)}$ across QPC_{L(R)}. For $1 < \tau_R \leq 2$, there are two channels across QPC_R—one fully ballistic and the other with transmission probability $\tau_R - 1$. With this approach, we achieve an accuracy down to 0.1% near the ballistic limit.

The sample is immersed into a perpendicular magnetic field $B \approx 4 \text{ T}$, which corresponds to the integer quantum Hall effect at filling

¹Centre de Nanosciences et de Nanotechnologies (C2N), CNRS, Université Paris Sud–Université Paris-Saclay, Université Paris Diderot–Sorbonne Paris Cité, 91120 Palaiseau, France. ²Institute for Theoretical Physics, ETH Zurich, CH-8093 Zurich, Switzerland. ³Département de Physique Théorique, Université de Genève, CH-1211 Genève, Switzerland. ⁴Department of Physics, Yale University, New Haven, Connecticut 06520, USA.

*These authors contributed equally to this work.

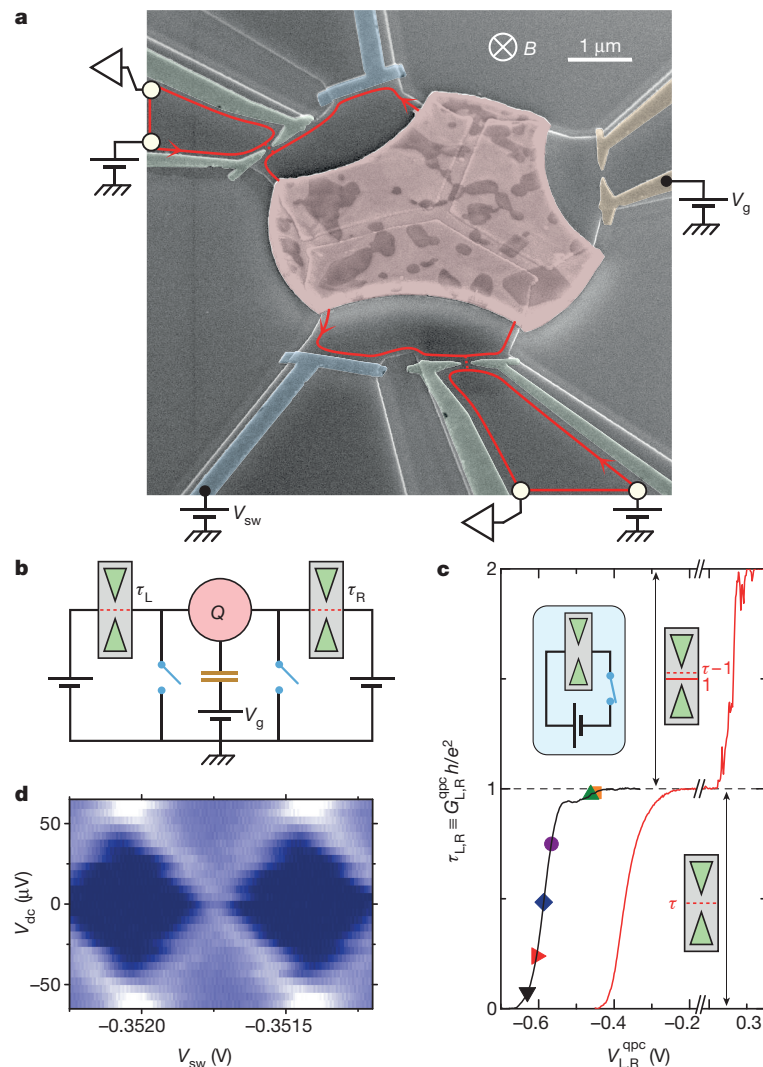


Figure 1 | Tunable quantum connection to a metallic island. **a**, Coloured sample micrograph. A micrometre-scale metallic island (red) is connected to large electrodes (white circles) through two quantum point contacts (QPCs, green split gates) formed in a buried two-dimensional electron gas (2DEG; darker grey delimited by bright lines). The lateral gates (blue) implement short-circuit switches as shown in **b**. The yellow gates, tuned at V_g negative enough to deplete the 2DEG underneath, are capacitively coupled to the island and used to evidence charge quantization. In the applied field $B \approx 4$ T, the current propagates along two edge channels (red lines) in the direction indicated by arrows. **b**, Sample schematic; colours as in **a**; Q represents the excess charge that can accumulate on the

metallic island. **c**, The ‘intrinsic’ (switch closed; see inset schematic) conductance $G_{L,R}^{qpc}$ across QPC_{L,R} (shown top-right and bottom left, respectively, in **a**) is shown versus split gate voltage $V_{L,R}^{qpc}$ as black (L) and red (R) lines. Symbols indicate the set-points of QPC_L used thereafter. The number and transmission probabilities of electronic channels through the QPC (pair of green triangles) are schematized for $\tau_R < 1$ and $\tau_R > 1$: a dashed (solid) red line represents a partially (perfectly) transmitted channel. **d**, Coulomb diamond patterns in the device conductance G_{SET} (larger shown brighter, from 0 in dark blue up to $0.13e^2/h$ in white) measured versus gate (V_{sw}) and bias (V_{dc}) voltages for tunnel contacts $\tau_{L,R} \ll 1$.

factor $\nu = 2$. In this regime, the electrical current propagates along two edge channels (shown as a single red line in Fig. 1a) in the direction indicated by arrows, which does not influence charge quantization (for a specific discussion see Methods section ‘Conductance in the near-ballistic regime with strong thermal fluctuations’). The large Zeeman splitting results in the full separation between the successive openings of the two spin-polarized quantum channels across the QPCs (Fig. 1c).

Charge quantization in the central island is unequivocally evidenced by periodic oscillations of the SET differential conductance G_{SET} (across QPC_L–island–QPC_R) when sweeping a capacitively coupled gate voltage, which develop into Coulomb diamonds with d.c. bias voltage V_{dc} (Fig. 1d). With both QPCs in the tunnel regime, $\tau_{L,R} \ll 1$, the span of the diamonds in V_{dc} gives the charging energy $E_C \approx k_B \times 0.3$ K ($C \approx 3.1$ fF).

We first probe the evolution of charge quantization with transmission probability directly from G_{SET} raw periodic modulations. Figure 2a

displays G_{SET} measured at $T \approx 17$ mK and $V_{dc} = 0$ while sweeping the capacitively coupled gate voltage V_g (Fig. 1a), for QPC_L fixed to $\tau_L = 0.24$ and with each panel corresponding to a different QPC_R tuning ($\tau_R = 0.1, 0.6, 0.88, 0.98$ and 1.5 , from left to right). These raw data reveal the remarkable robustness of charge quantization to connection strength. At $\tau_R = 0.1$ and $\tau_R = 0.6$, the presence of sharp periodic peaks separated by intervals in which $G_{SET} \approx 0$ signals an essentially unaltered charge quantization over the greater part of transmission probabilities. Although $G_{SET}(\delta V_g)$ progressively evolves with increasing $\tau_R < 1$ into a sinusoid with non-zero minima, relatively important modulations of fixed (τ_R -independent) period persist very close to the ballistic limit, at $\tau_R = 0.98$. In stark contrast, G_{SET} is independent of V_g at $\tau_R = 1.5$, confirming the predicted complete collapse of charge quantization in the presence of a fully ballistic channel. Note that G_{SET} remains reduced by Coulomb interactions, even at $\tau_R = 1.5$, as evidenced by the pronounced conductance dip at low V_{dc} (inset of Fig. 2b).

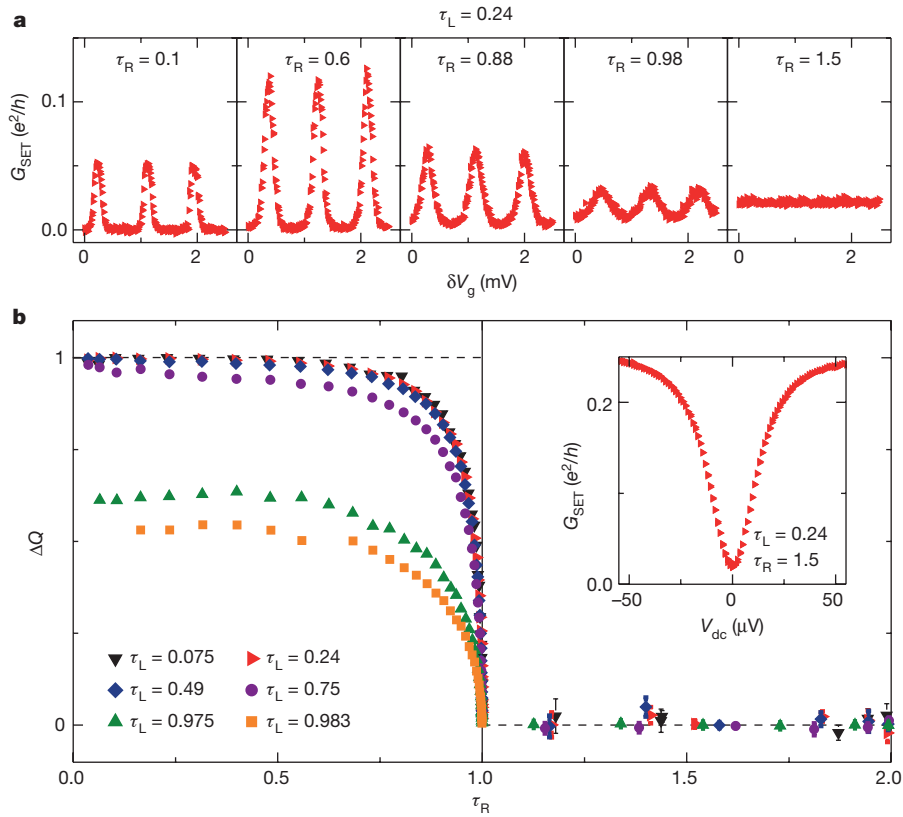


Figure 2 | Charge quantization versus connection strength at $T \approx 17$ mK. **a**, Conductance sweeps $G_{\text{SET}}(\delta V_g)$ with fixed $\tau_L = 0.24$ and varying $\tau_R = 0.1, 0.6, 0.88, 0.98$ and 1.5 , from left to right, as indicated. **b**, Visibility of G_{SET} oscillations $\Delta Q \equiv (G_{\text{SET}}^{\text{max}} - G_{\text{SET}}^{\text{min}})/(G_{\text{SET}}^{\text{max}} + G_{\text{SET}}^{\text{min}})$ versus τ_R , with each set of symbols corresponding to a different QPC_L set-point τ_L , as indicated, corresponding to those indicated by the matching symbols

in Fig. 1c. Inset, dynamical Coulomb blockade renormalization of G_{SET} versus d.c. voltage V_{dc} in the absence of charge quantization, at $\tau_L = 0.24$ and $\tau_R = 1.5$. The error bars are the standard error on the mean value of ΔQ , obtained from the statistical uncertainty of about ten measurements of G_{SET} (see Methods).

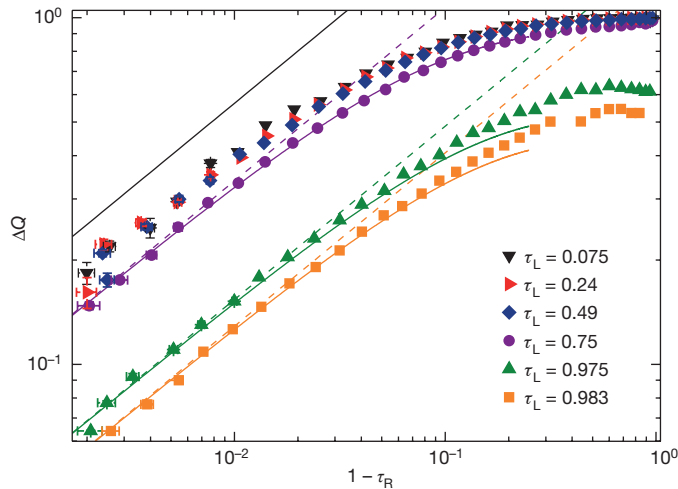


Figure 3 | Charge quantization scaling near the ballistic critical point. The ΔQ data at $T \approx 17$ mK are displayed versus $1 - \tau_R$ on a log-log scale, with different symbols for the different QPC_L set-points, as in Fig. 2. Solid lines are quantitative predictions (no fit parameters) derived assuming $k_B T \ll E_C$, $1 - \tau_R \ll 1$ and either $\tau_L \ll 1$ (top (black) line) or $1 - \tau_L \ll 1$ (bottom three (purple, green and orange) lines). The power law $\Delta Q \propto \sqrt{1 - \tau_R}$ (straight, dashed lines) is systematically observed for $1 - \tau_R \lesssim 0.02$ and at intermediate τ_L . The horizontal error bars arise from the dispersion of at least 40 transmission settings; the vertical error bars are calculated from the statistical uncertainty of about 10 measurements of one period of G_{SET} (see Methods).

Indeed, the so-called dynamical Coulomb blockade does not rely on a quantized island charge, but results from the discreteness of charge transfers across non-ballistic channels^{1,2}.

The degree of charge quantization versus connection strength is characterized, separately from the dynamical Coulomb blockade renormalization of the channels, by focusing on the visibility of the periodic modulations $\Delta Q \equiv (G_{\text{SET}}^{\text{max}} - G_{\text{SET}}^{\text{min}})/(G_{\text{SET}}^{\text{max}} + G_{\text{SET}}^{\text{min}})$, with $G_{\text{SET}}^{\text{max}(\text{min})}$ the maximum (minimum) SET conductance over one gate-voltage period and, from now on, $V_{\text{dc}} = 0$. A visibility $\Delta Q = 1$ ($\Delta Q = 0$) signals a full (an absence of) charge quantization. Moreover, the visibility ΔQ is directly proportional to the charge oscillations of the island with gate voltage (that is, charge quantization) when one channel approaches the ballistic limit (for example, $\tau_R \rightarrow 1$)^{7,25–27}. As put forward in ref. 26, this proportionality coefficient reduces to the numerical factor $e/(2\pi \times 1.59)$ for $\tau_L \ll 1$ and $k_B T \ll E_C$.

Figure 2b shows ΔQ versus τ_R at $T \approx 17$ mK, with each set of symbols corresponding to a different tuning of the second QPC ($\tau_L \in \{0.075, 0.24, 0.49, 0.75, 0.975, 0.983\}$). The robustness of charge quantization with the connection strength of one channel (τ_R) is established independently of the second channel (τ_L), from the nearly constant ΔQ for $\tau_R \lesssim 0.6$. When further increasing τ_R , ΔQ noticeably diminishes and systematically collapses to zero precisely at the ballistic critical point $\tau_R = 1$. For $\tau_R \geq 1$, in the presence of one ballistic channel, ΔQ remains perfectly null at experimental accuracy (see Methods for additional tests).

Power laws characterizing the scaling of charge quantization as $\tau_R \rightarrow 1$ are best revealed by plotting ΔQ versus the ‘distance’ from the ballistic critical point $1 - \tau_R > 0$ on a log-log scale. As shown in Fig. 3, the $T = 17$ mK data (symbols) systematically vanish as $\sqrt{1 - \tau_R}$ (straight lines) for $1 - \tau_R \lesssim 0.02$.

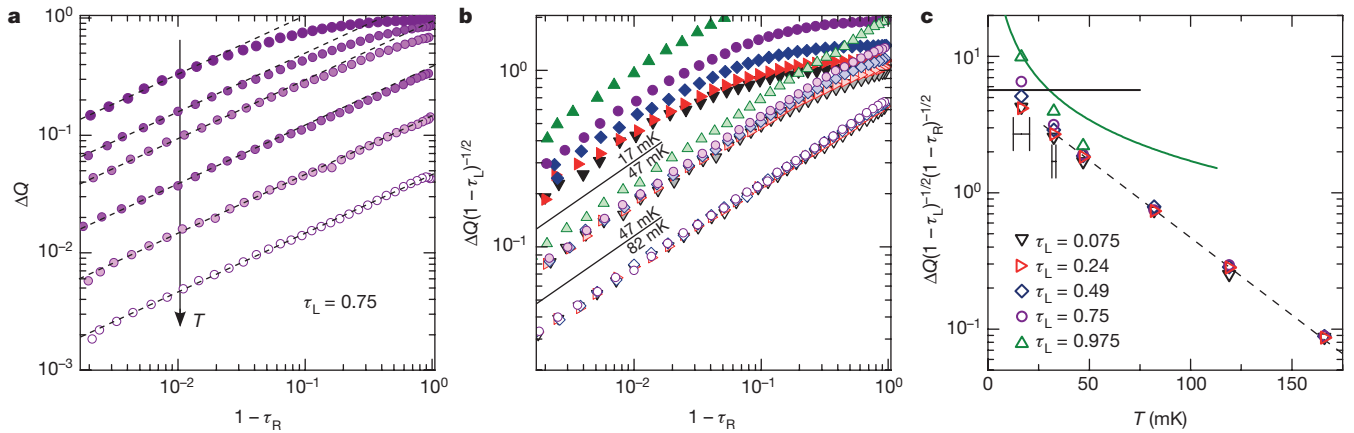


Figure 4 | Crossover to a universal charge quantization scaling as temperature is increased. **a**, Symbols display ΔQ versus $1 - \tau_R$ at $\tau_L = 0.75$ and for $T \approx 17$ mK, 32 mK, 47 mK, 82 mK, 119 mK and 166 mK, from top to bottom. The τ_R range over which $\Delta Q \propto \sqrt{1 - \tau_R}$ (straight, dashed lines) extends up to the full interval $\tau_R \in [0, 1]$ when increasing T . **b**, The rescaled $\Delta Q/\sqrt{1 - \tau_L}$ is shown versus $1 - \tau_R$, with a different set of symbols corresponding to different QPC_L set-points as in **c**. Solid lines separate the data at $T \approx 17$ mK (top, darker filling), $T \approx 47$ mK (middle) and $T \approx 82$ mK (bottom, brighter filling). At $T = 82$ mK, all the data collapse onto a single universal curve $\Delta Q \propto \sqrt{(1 - \tau_L)(1 - \tau_R)}$. Symbols

display the fully rescaled data $\Delta Q/\sqrt{(1 - \tau_L)(1 - \tau_R)}$ versus T on semi-log scale, extracted in the regime in which $1 - \tau_R$ is small enough that $\Delta Q \propto \sqrt{1 - \tau_R}$; data for $\tau_L = 0.975$ are plotted only for $T \leq 47$ mK. Horizontal error bars represent the experimental temperature uncertainty at $T = 17 \pm 4$ mK and $T = 32 \pm 1$ mK. Solid lines are the quantitative predictions in the quantum regime $k_B T \ll E_C$, given by equation (1) (black, horizontal) and equation (2) (green, curved). The straight dashed line displays an exponential decay close to predictions in the presence of strong thermal fluctuations (see text).

The Coulomb blockade theory of electronic transport in the presence of a nearly ballistic channel ($1 - \tau_R \ll 1$) relies on the bosonization approach that was initially developed to address correlated electrons in one dimension. Quantitative predictions were obtained for $k_B T \ll E_C$ and for a second channel in either the tunnel ($\tau_L \ll 1$) or almost-ballistic ($1 - \tau_L \ll 1$) regime^{25,28}. In both cases, ΔQ is expected to vanish as $\sqrt{1 - \tau_R}$:

$$\Delta Q(1 - \tau_R \ll 1; \tau_L \ll 1, k_B T \ll E_C) \approx 5.7 \sqrt{1 - \tau_R} \quad (1)$$

$$\Delta Q \left(\sqrt{1 - \tau_{L,R}} \ll \frac{k_B T}{E_C} \ll 1 \right) \approx \frac{0.57 E_C}{k_B T} \sqrt{(1 - \tau_L)(1 - \tau_R)} \quad (2)$$

Such a scaling, initially proposed in ref. 6, was also predicted for the gate-voltage modulation of thermodynamic quantities for multi-channel junctions using an extension⁸ of the instanton technique^{1,29}.

The data establishes the $\sqrt{1 - \tau_R}$ scaling for arbitrary $\tau_L \in [0, 1]$, beyond the tunnel and ballistic limits currently accessible to transport theory. The dashed lines in Fig. 3 display the asymptotic ($\sqrt{1 - \tau_{L,R}} \ll k_B T/E_C$), quantitative predictions of equation (2) for our completely characterized device at $T = 17$ mK, without fitting parameters. The non-asymptotic ΔQ predictions (equation (1) for $\tau_L \ll 1$; see Methods for $1 - \tau_L \ll 1$) are shown versus $1 - \tau_R < 0.25$ as solid lines. Data and quantitative predictions are indistinguishable for $1 - \tau_R \lesssim 0.1$ for $\tau_L = 0.983$, $\tau_L = 0.975$ and, more surprisingly, $\tau_L = 0.75$. The equation (1) prediction (black line in Fig. 3) remains noticeably (about 25%) above the $\tau_L = 0.075$ data for $1 - \tau_R \ll 1$. This numerical difference could result from the finite experimental T , because equation (1) is exact only at $T = 0$.

We now investigate the ways in which the combination of thermal and quantum fluctuations impacts the quantization of charge. As temperature rises, the population of additional charge states is expected to average out charge quantization^{1,2}. Figure 4a displays the measured ΔQ (symbols) versus $1 - \tau_R$ at different temperatures, from $T = 17$ mK (darker filling) to $T = 166$ mK (brighter filling), for the representative QPC_L setting $\tau_L = 0.75$. As naively expected, ΔQ decreases as T increases. In line with thermodynamic expectations⁸ (Methods), the $\Delta Q \propto \sqrt{1 - \tau_R}$ scaling (straight lines) that originates from quantum

fluctuations not only persists for increasing T , but extends over a widening range of τ_R , up to the full-scale $\tau_R \in [0, 1]$.

The crossover towards this universal behaviour is established by comparing the rescaled visibility $\Delta Q/\sqrt{1 - \tau_L}$ for different τ_L settings with $1 - \tau_R$. The symbols in Fig. 4b represent the rescaled data at $T = 17$ mK, $T = 47$ mK and $T = 82$ mK, with brighter filling at higher temperatures. As T increases, the scatter associated with the various τ_L values narrows. For $T \geq 82$ mK, the rescaled data collapse onto a single, universal (for all τ_L), straight line $\Delta Q \propto \sqrt{(1 - \tau_L)(1 - \tau_R)}$ over the full range $\tau_{L,R} \in [0, 1]$.

The temperature dependence is further characterized by plotting $\Delta Q/\sqrt{(1 - \tau_L)(1 - \tau_R)}$ (determined at low enough $1 - \tau_R$ such that $\Delta Q \propto \sqrt{1 - \tau_R}$) versus temperature on a semi-log scale (Fig. 4c, symbols). The $k_B T \ll E_C$ prediction of equation (1) (equation (2)) is displayed as a black (green) solid line for $T < 75$ mK ($T < 115$ mK). We find for $T \geq 82$ mK (up to 166 mK, $2.8 \leq \pi^2 k_B T/E_C \leq 5.6$) that the different τ_L data points collapse onto the same exponential decay (dashed line in Fig. 4c): $\Delta Q \approx \sqrt{(1 - \tau_L)(1 - \tau_R)} \exp(-0.8 \pi^2 k_B T/E_C)$. We have extended the Coulomb blockade theory for the conductance to include thermal fluctuations in the limits of tunnel or nearly ballistic channels (Methods). In the regime of strong thermal averaging, we predict $\Delta Q \propto \sqrt{(1 - \tau_L)(1 - \tau_R)} \exp(-\pi^2 k_B T/E_C)$ (neglecting factors not exponential in T)—a dependence that is also expected for thermodynamic properties⁸ (Methods)—in close agreement with the experimental findings regarding the effect of $\tau_{L,R}$ and T .

Although theoretical predictions for low-temperature transport currently apply to only the nearly ballistic and tunnel limits, we anticipate that recent advances, including those in numerical renormalization group methods³⁰, will open up access to the full range of connection strengths. Our results may therefore provide a test-bed for strongly correlated electron-theoretical methods, for which non-perturbative techniques are ubiquitous. The understanding and on-demand control of charge quantization in mesoscopic circuits might lead to applications beyond the field of single electronics. The central role of charge quantization in the different quantum laws of electricity with coherent conductors indicates that direct quantum engineering could have implications for future nanoelectronics, such as semiconductor-metal hybrid devices that are crucial for developing topologically protected quantum bits⁹. The hybrid implementation we have presented also enables further

fundamental exploration, including of charge quantization with correlated electrons such as in the multi-channel Kondo regime and/or with fractionally charged anyonic quasiparticles.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 March; accepted 10 June 2016.

- Schön, G. & Zaikin, A. D. Quantum coherent effects, phase transitions, and the dissipative dynamics of ultra small tunnel junctions. *Phys. Rep.* **198**, 237–412 (1990).
- Grabert, H. & Devoret, M. H. (eds). *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures* (Plenum, 1992).
- Likharev, K. K. Single-electron devices and their applications. *Proc. IEEE* **87**, 606–632 (1999).
- Meschke, M., Engert, J., Heyer, D. & Pekola, J. P. Comparison of Coulomb blockade thermometers with the International Temperature Scale PLTS-2000. *Int. J. Thermophys.* **32**, 1378–1386 (2011).
- Pekola, J. P. et al. Single-electron current sources: toward a refined definition of the ampere. *Rev. Mod. Phys.* **85**, 1421–1472 (2013).
- Flensberg, K. Capacitance and conductance of mesoscopic systems connected by quantum point contacts. *Phys. Rev. B* **48**, 11156–11166 (1993).
- Matveev, K. A. Coulomb blockade at almost perfect transmission. *Phys. Rev. B* **51**, 1743–1751 (1995).
- Nazarov, Y. V. Coulomb blockade without tunnel junctions. *Phys. Rev. Lett.* **82**, 1245–1248 (1999).
- Albrecht, S. M. et al. Exponential protection of zero modes in Majorana islands. *Nature* **531**, 206–209 (2016).
- Kouwenhoven, L. et al. Single electron charging effects in semiconductor quantum dots. *Z. Phys. B* **85**, 367–373 (1991).
- Staring, A. A. M., Williamson, J. G., van Houten, H. & Beenakker, C. W. J. Coulomb-blockade oscillations in a quantum dot. *Physica B* **175**, 226–230 (1991).
- van der Vaart, N. et al. Charging effects in quantum dots at high magnetic fields. *Physica B* **189**, 99–110 (1993).
- Molenkamp, L. W., Flensberg, K. & Kemerink, M. Scaling of the Coulomb energy due to quantum fluctuations in the charge on a quantum dot. *Phys. Rev. Lett.* **75**, 4282–4285 (1995).
- Joyez, P., Bouchiat, V., Esteve, D., Urbina, C. & Devoret, M. H. Strong tunneling in the single-electron transistor. *Phys. Rev. Lett.* **79**, 1349–1352 (1997).
- Chouvaev, D., Kuzmin, L. S., Golubev, D. S. & Zaikin, A. D. Strong tunneling and Coulomb blockade in a single-electron transistor. *Phys. Rev. B* **59**, 10599–10602 (1999).
- Berman, D., Zhitenev, N. B., Ashoori, R. C. & Shayegan, M. Observation of quantum fluctuations of charge on a quantum dot. *Phys. Rev. Lett.* **82**, 161–164 (1999).
- Duncan, D. S., Livermore, C., Westervelt, R. M., Maranowski, K. D. & Gossard, A. C. Direct measurement of the destruction of charge quantization in a single-electron box. *Appl. Phys. Lett.* **74**, 1045–1047 (1999).
- Amasha, S. et al. Coulomb blockade in an open quantum dot. *Phys. Rev. Lett.* **107**, 216804 (2011).
- Pasquier, C. et al. Quantum limitation on Coulomb blockade observed in a 2D electron system. *Phys. Rev. Lett.* **70**, 69–72 (1993).
- Crouch, C. H., Livermore, C., Westervelt, R. M., Campman, K. L. & Gossard, A. C. Coulomb oscillations in partially open quantum dots. *Superlattices Microstruct.* **20**, 377–381 (1996).
- Liang, C.-T. et al. Experimental evidence for Coulomb charging effects in an open quantum dot at zero magnetic field. *Phys. Rev. Lett.* **81**, 3507–3510 (1998).
- Cronenwett, S. M. et al. Mesoscopic Coulomb blockade in one-channel quantum dots. *Phys. Rev. Lett.* **81**, 5904–5907 (1998).
- Tkachenko, O. A. et al. Coulomb charging effects in an open quantum dot device. *J. Phys. Condens. Matter* **13**, 9515–9534 (2001).
- Aleiner, I. L. & Glazman, L. I. Mesoscopic charge quantization. *Phys. Rev. B* **57**, 9608–9641 (1998).
- Furusaki, A. & Matveev, K. A. Theory of strong inelastic cotunneling. *Phys. Rev. B* **52**, 16676–16695 (1995).
- Yi, H. & Kane, C. L. Coulomb blockade in a quantum dot coupled strongly to a lead. *Phys. Rev. B* **53**, 12956–12966 (1996).
- Le Hur, K. & Seelig, G. Capacitance of a quantum dot from the channel-anisotropic two-channel Kondo model. *Phys. Rev. B* **65**, 165338 (2002).
- Matveev, K. A. & Andreev, A. V. Thermopower of a single-electron transistor in the regime of strong inelastic cotunneling. *Phys. Rev. B* **66**, 045301 (2002).
- Korshunov, S. E. Coherent and incoherent tunneling in a Josephson junction with a “periodic” dissipation. *JETP Lett.* **45**, 434–436 (1987).
- Mitchell, A. K., Landau, L. A., Fritz, L. & Sela, E. Universality and scaling in a charge two-channel Kondo device. *Phys. Rev. Lett.* **116**, 157202 (2016).

Acknowledgements This work was supported by the European Research Council (ERC-2010-StG-20091028, no. 259033), the French RENATECH network, the national French programme ‘Investissements d’Avenir’ (Labex NanoSaclay, ANR-10-LABX-0035), the US Department of Energy (DE-FG02-08ER46482) and the Swiss National Science Foundation.

Author Contributions S.J. and Z.I. performed the experiment with inputs from A.A. and F.P.; S.J., Z.I., A.A. and F.P. analysed the data; F.D.P. fabricated the sample and contributed to a preliminary experiment; U.G., A.C. and A.O. grew the 2DEG; I.P.L., E.I., E.V.S. and L.I.G. developed the strong thermal fluctuations theory; F.P. led the project and wrote the manuscript with inputs from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.P. (frederic.pierre@lpn.cnrs.fr).

Reviewer Information *Nature* thanks Y. Nazarov and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Sample. The sample is nanostructured by standard e-beam lithography in a GaAs/Ga(Al)As 2DEG located 105 nm below the surface, of density $2.5 \times 10^{11} \text{ cm}^{-2}$ and mobility $10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. The ohmic contact between the micrometre-scale metallic island and the buried 2DEG is obtained by thermal diffusion into the semiconductor of a metallic multilayer of nickel (30 nm), gold (120 nm) and germanium (60 nm); see, for example, ref. 31. See methods in ref. 32 for the estimation of the typical energy spacing between electronic levels in the central metallic island on the same sample.

Experimental set-up. The measurements were performed in a dilution refrigerator including multiple filters along the electrical lines and two shields at the mixing chamber. Conductance measurements were carried out by standard lock-in techniques at low frequencies, below 100 Hz, taking advantage of the chiral current propagation in the quantum Hall regime (see Extended Data Fig. 1). Noise measurements for the electronic temperature were performed in the megahertz range using a homemade cryogenic amplifier (for details, see the supplementary information of ref. 33).

Electronic temperature. The displayed electronic temperatures correspond to those extracted on-chip using either quantum shot noise primary thermometry³⁴ or thermal noise thermometry, with error bars encapsulating also the outcome of Coulomb blockade oscillations primary thermometry (at $T \leq 32 \text{ mK}$) and/or standard thermometry from RuO_2 resistors thermally anchored to the mixing chamber (at $T \geq 32 \text{ mK}$).

Interface between the metallic island and the 2DEG. A 2DEG-metallic island transmission probability $\tau_{\Omega\text{-out}} > 0.9995$ is obtained with the self-calibrated procedure described below. Here, the switches are set in open positions as in Fig. 1b (with edge channels following the red lines shown Fig. 1a and Extended Data Fig. 1). First, $\text{QPC}_{\text{L,R}}$ are set at $\tau_{\text{L,R}} = 1$, in the middle of the very flat and broad intermediate plateau (owing to the robust quantum Hall effect), and we measure the reflected signal $V_{\text{RR}}^{\tau_{\text{L,R}}=1}$ (see Extended Data Fig. 1). The average transmission probability $\tau_{\Omega\text{-out}}$ of the first (outer-edge quantum Hall) channel emitted from QPC_{L} and QPC_{R} into the metallic island then reads:

$$V_{\text{RR}}^{\tau_{\text{L,R}}=1} = G_{\text{R}}(1 - \tau_{\Omega\text{-out}}/4)V_{\text{R}}$$

with V_{R} the (a.c.) voltage applied at the input of QPC_{R} (see Extended Data Fig. 1) and G_{R} the gain of amplification chain R. Second, we eliminate calibration uncertainties by measuring the reflected signal $V_{\text{RR}}^{\tau_{\text{L,R}}=0} = G_{\text{R}}V_{\text{R}}$ with $\text{QPC}_{\text{L,R}}$ depleted ($\tau_{\text{L,R}} = 0$). The ratio $V_{\text{RR}}^{\tau_{\text{L,R}}=1}/V_{\text{RR}}^{\tau_{\text{L,R}}=0}$ gives $\tau_{\Omega\text{-out}}$ directly. With this approach, we obtain $|1 - \tau_{\Omega\text{-out}}| < 5 \times 10^{-4}$ ($\tau_{\Omega\text{-out}} \approx 0.9997 \pm 0.0002$). The same approach including also the second (inner-edge quantum Hall) channel gives $\tau_{\Omega\text{-in}} \approx 0.9976$. Note that it is usual to have better ohmic contacts with the outer quantum Hall channels, which are closest to the sample edges.

Short-circuit switch operation. In practice, closing the short-circuit switches is realized by changing the voltage applied to the adjacent characterization gate (blue in Fig. 1a, see Extended Data Fig. 2a for the conductance versus gate voltage of switch R) from -0.35 V (2DEG depleted/switch open) to 0.1 V (two edge channels perfectly transmitted/switch closed).

Capacitive crosstalk corrections. The transmission probability across each QPC is slightly modified when changing the voltage applied either to its adjacent characterization gate or to the gate tuning the other QPC. Owing to the large, micrometre-scale distances, this modification remains relatively small, particularly near the ballistic critical point ($< 1\%$ for $\tau_{\text{L,R}} \in [0.9, 1]$ when changing the adjacent switch from closed to open). Let us first consider the crosstalk from one QPC to the other, which is more straightforward to extract. For this purpose, the characterization gate adjacent to the QPC, for which the crosstalk is to be compensated, is set to its short-circuit/closed position (as in Fig. 1c) such that the gate-voltage change that tunes the other QPC is felt only through the capacitive crosstalk. We find that this crosstalk can be precisely compensated by a relatively small shift (approximately -1%) of the split gate voltage. Regarding the capacitive crosstalk due to the adjacent characterization gate, the difficulty is to isolate this contribution from changes in the Coulomb blockade renormalization of the QPC conductance. To suppress this renormalization, the other QPC is set in the middle of its $\tau_{\text{L(R)}} = 1$ plateau and we apply a large d.c. bias voltage compared to the charging energy. Extended Data Fig. 2b displays the differential conductance of QPC_{R} , measured in the presence of the applied bias $V_{\text{R}} = 72 \mu\text{V}_{\text{dc}}$, versus gate voltage $V_{\text{R}}^{\text{qpc}}$ for the adjacent switch set to position open (red line) and closed (blue line). The gate voltage shift $\Delta V_{\text{R}}^{\text{ct}}$ that is needed to compensate the crosstalk is determined at low QPC conductances $G_{\text{R}}^{\text{qpc}} \lesssim 0.1 e^2/h$, for which the d.c. voltage drop across the QPC is nearly independent of the switch position. Extended Data Fig. 2c displays as symbols the crosstalk compensation for QPC_{R} in response to increasing the adjacent characterization gate voltage from $V_{\text{R}}^{\text{sw}} = -0.5 \text{ V}$. The amplitude of the negative crosstalk compensation is found to increase linearly, with different slopes

for different values of the switch conductance G_{R}^{sw} . Indeed, the capacitive crosstalk depends on the precise paths of the edge channels, which screen the gate potentials. The crosstalk compensations used in the experiment when setting the adjacent switch from open to closed are $\Delta V_{\text{R}}^{\text{qpc}} \approx -6 \text{ mV}$ for QPC_{R} and $\Delta V_{\text{L}}^{\text{qpc}} \approx -10 \text{ mV}$ for QPC_{L} .

Calibrations. The reflected signal V_{RR} is normalized by the signal $V_{\text{RR}}^{\tau_{\text{L,R}}=0}$ measured when setting $\tau_{\text{L,R}} = 0$. The injection voltage and amplifier gain thereby cancel out in the expression of the SET conductance G_{SET} :

$$G_{\text{SET}} = \frac{2e^2}{h}(1 - V_{\text{RR}}/V_{\text{RR}}^{\tau_{\text{L,R}}=0})$$

To reduce the noise level, we also extract G_{SET} from the (redundant) transmitted signal V_{LR} (see Extended Data Fig. 1):

$$G_{\text{SET}} = \frac{2e^2}{h}(V_{\text{LR}}/V_{\text{RR}}^{\tau_{\text{L,R}}=0})G_{\text{R}}/G_{\text{L}}$$

with G_{R} (G_{L}) the gain of amplification chain R (L). The ratio $G_{\text{R}}/G_{\text{L}}$ is determined by setting $\text{QPC}_{\text{L,R}}$ at $\tau_{\text{L,R}} = 1$ and measuring both the signals reflected ($V_{\text{RR}}^{\tau_{\text{L,R}}=1}$) and transmitted ($V_{\text{LR}}^{\tau_{\text{L,R}}=1}$):

$$\frac{G_{\text{R}}}{G_{\text{L}}} = \frac{1 - V_{\text{RR}}^{\tau_{\text{L,R}}=1}/V_{\text{RR}}^{\tau_{\text{L,R}}=0}}{V_{\text{LR}}^{\tau_{\text{L,R}}=1}/V_{\text{RR}}^{\tau_{\text{L,R}}=0}} \approx 1.0105$$

Experimental determination of ΔQ . For $\tau_{\text{R}} \leq 0.99$, the signal-to-noise ratio is always sufficient to accurately extract the values of $G_{\text{SET}}^{\text{max,min}}$ directly from the periodic conductance maximums and minimums, which stand out very strongly from the background noise. The error bars on the visibility $\Delta Q \equiv (G_{\text{SET}}^{\text{max}} - G_{\text{SET}}^{\text{min}})/(G_{\text{SET}}^{\text{max}} + G_{\text{SET}}^{\text{min}})$ were calculated from the statistical uncertainty on $G_{\text{SET}}^{\text{max,min}}$, which is typically estimated from ten different sweeps of one period. In this regime ($\tau_{\text{R}} \leq 0.99$), the calculated error bars are smaller than the size of the symbol and are therefore not shown.

For $\tau_{\text{R}} \in [0.99, 0.998]$, although the periodic oscillations can still be clearly distinguished in the raw data (see Extended Data Fig. 3), the above direct procedure would result in uncertainties that can become quite large, especially at base temperature and in the presence of a weakly transmitted second channel ($\tau_{\text{L}} = 0.075$). To improve our extraction of ΔQ , we take advantage of the observation that the conductance oscillations are sinusoidal for $\tau_{\text{R}} \geq 0.98$ (see Extended Data Fig. 3), as is expected from theory (see equations (3) and (6), and solid lines in Extended Data Fig. 3): the visibility of the conductance oscillations ΔQ is then extracted from a sinusoidal fit of the conductance sweeps $G_{\text{SET}}(V_{\text{g}})$. The displayed error bars are the statistical error on the mean value obtained from the distinct ΔQ values obtained by separately fitting approximately six different conductance sweeps. The two procedures give the same value of ΔQ in the intermediate regime $\tau_{\text{R}} \in [0.98, 0.99]$ for which they both accurately apply.

For $\tau_{\text{R}} \geq 1$, there are no periodic oscillations directly visible in the raw conductance sweeps $G_{\text{SET}}(V_{\text{g}})$ (see right panel in Fig. 2a). To put experimental bounds on the basic statement $\Delta Q \approx 0$, we determined the visibility ΔQ (displayed Fig. 2b) using the following procedure. First, we determine the most probable positions of the conductance maximums and minimums by ‘fitting’ a conductance sweep (typically extending over ten Coulomb oscillation periods) with a sinusoidal function at the known period of Coulomb oscillations, using its phase as a fitting parameter. For each of these positions, a different value of $G_{\text{SET}}^{\text{max}}$ or $G_{\text{SET}}^{\text{min}}$ is obtained by averaging the data over an extension of one quarter of a period (assuming sinusoidal oscillations, this would result in a visibility reduction smaller than 10%). By separately extracting $G_{\text{SET}}^{\text{max,min}}$ for the approximately ten periods, we calculate their mean values and estimate the corresponding standard errors. The error bars displayed Fig. 2b are the standard error on the mean value of ΔQ , obtained from the statistical uncertainty on $G_{\text{SET}}^{\text{max,min}}$.

Predictions in the quantum asymmetric regime. In the quantum asymmetric regime ($k_{\text{B}}T \ll E_{\text{C}}$, $\tau_{\text{L}} \ll 1$, $1 - \tau_{\text{L}} \ll 1$), the conductance reads (equation (34) in ref. 28):

$$G_{\text{SET}}^{\tau_{\text{L}} \ll 1, 1 - \tau_{\text{R}} \ll 1} = \tau_{\text{L}} \frac{e^2}{h} \frac{2\pi^4 (k_{\text{B}}T)^2}{3\gamma^2 E_{\text{C}}} \times [1 - 2\gamma\xi\sqrt{1 - \tau_{\text{R}}}\cos(2\pi\delta V_{\text{g}}/\Delta)] \quad (3)$$

with $\gamma \approx \exp(0.5772)$, $\xi \approx 1.59$, Δ the gate-voltage period and δV_{g} the gate-voltage difference from charge degeneracy. In the ballistic limit ($1 - \tau_{\text{R}} = 0$), the conductance does not depend on gate voltage, but vanishes as T^2 following quantitatively, with the exact same pre-factor, the dynamical Coulomb blockade predictions² for the same E_{C} and the corresponding series resistance $R = h/e^2$. Using equation (3), the visibility of the oscillations of conductance reads:

$$\Delta Q(\tau_{\text{L}} \ll 1, 1 - \tau_{\text{R}} \ll 1) = 2\gamma\xi\sqrt{1 - \tau_{\text{R}}} \quad (4)$$

The temperature dependence of $G_{\text{SET}}^{\tau_L \ll 1, 1 - \tau_R \ll 1}$ (associated with dynamical Coulomb blockade) cancels out in ΔQ . Charge discreteness also affects the gate-voltage dependence of thermodynamic quantities, such as the average charge $\langle Q \rangle$ or the differential capacitance ($C_{\text{diff}} \equiv \partial \langle Q \rangle / \partial V_g$). The effect of Coulomb blockade on thermodynamic quantities was studied most comprehensively for tunnel junctions^{1,29}: at $T = 0$ and $G \gg e^2/h$, the amplitude of average charge oscillations decays exponentially with Gh/e^2 ; see, for example, refs 35–37. The theoretical extension to multi-channel junctions of arbitrary transmission, beyond the tunnel limit, was performed in ref. 8. In the presence of a single, nearly ballistic channel, the bosonization approach allows for an exact solution of the average charge in the metallic island in the low-energy ‘quantum’ regime $k_B T \ll E_C$ (equation (26) in ref. 7):

$$\langle Q \rangle^{\tau_L \ll 1, 1 - \tau_R \ll 1} = \frac{eV_g}{\Delta} - \frac{e\gamma}{\pi} \sqrt{1 - \tau_R} \sin(2\pi V_g/\Delta) + Q_0$$

with Q_0 a charge offset. In the ballistic limit ($1 - \tau_R = 0$), the charge increases linearly with gate voltage, corresponding to an absence of charge quantization. The degree of charge quantization can be characterized by the relative amplitude of the oscillations of charge or, equivalently, by the visibility of the differential capacitance ($C_{\text{diff}} \equiv \partial \langle Q \rangle / \partial V_g$) oscillations:

$$\Delta C_{\text{diff}}(\tau_L \ll 1, 1 - \tau_R \ll 1) \equiv \frac{C_{\text{diff}}^{\text{max}} - C_{\text{diff}}^{\text{min}}}{C_{\text{diff}}^{\text{max}} + C_{\text{diff}}^{\text{min}}} = 2\gamma \sqrt{1 - \tau_R} \quad (5)$$

The degree of charge quantization vanishes as $\sqrt{1 - \tau_R}$ when approaching the ballistic limit, and does not depend on temperature in the quantum regime ($k_B T \ll E_C$). Importantly, the visibility in the SET conductance oscillations is directly proportional to the visibility of the differential capacitance oscillations²⁶, up to the fixed numerical factor $\xi \approx 1.59$:

$$\Delta Q(\tau_L \ll 1, 1 - \tau_R \ll 1) = \xi \Delta C_{\text{diff}}(\tau_L \ll 1, 1 - \tau_R \ll 1)$$

Predictions in the quantum near-ballistic regime. In the quantum near-ballistic regime ($k_B T \ll E_C$, $1 - \tau_{L,R} \ll 1$), the conductance G_{SET} reads (equations (38) and (26) in ref. 25):

$$G_{\text{SET}} = \frac{e^2}{2h} \left[1 - \int_0^\infty \frac{\Gamma_-^2 / \cosh^2(x)}{(x^2 \pi^2 k_B T / \gamma E_C)^2 + \Gamma_-^2} dx \right] \quad (6)$$

with $\gamma \approx \exp(0.5772)$ and

$$\Gamma_- = (1 - \tau_L) + (1 - \tau_R) - 2\sqrt{(1 - \tau_L)(1 - \tau_R)} \cos(2\pi \delta V_g / \Delta)$$

with Δ the gate-voltage period. The quantitative ΔQ predictions calculated with the maximum and minimum of G_{SET} inferred from equation (6) are displayed as coloured solid lines in Fig. 3. When approaching the ballistic critical point ($\sqrt{1 - \tau_{L,R}} \ll k_B T / E_C \ll 1$), the visibility ΔQ reduces to the simple asymptotic expression (equation (2) with $\gamma/\pi \approx 0.57$, reprinted here for convenience):

$$\Delta Q \left(\sqrt{1 - \tau_{L,R}} \ll \frac{k_B T}{E_C} \ll 1 \right) = \frac{\gamma E_C}{\pi k_B T} \sqrt{(1 - \tau_L)(1 - \tau_R)} \quad (7)$$

The differential capacitance (C_{diff}) when one QPC approaches the ballistic critical point ($\tau_R \rightarrow 1$) reduces to the asymptotic expression (equation (41) in ref. 27):

$$C_{\text{diff}}^{1 - \tau_R \ll 1 - \tau_L \ll 1} = -4\gamma \frac{e}{\Delta} \ln(1 - \tau_L) \sqrt{(1 - \tau_L)(1 - \tau_R)} \cos\left(\frac{2\pi \delta V_g}{\Delta}\right) + \frac{e}{\Delta}$$

and the visibility in the oscillations of the differential capacitance reads:

$$\Delta C_{\text{diff}}(1 - \tau_R \ll 1 - \tau_L \ll 1) = -4\gamma \ln(1 - \tau_L) \sqrt{(1 - \tau_L)(1 - \tau_R)}$$

We recover the same $\sqrt{1 - \tau_R}$ scaling behaviour near the ballistic critical point ($\tau_R = 1$) that was found in the asymmetric regime (equations (4) and (5)), and which is also found in the visibility of the conductance Coulomb oscillations (equation (7)). For two identical (for example, spin-degenerate) channels ($\tau \equiv \tau_L = \tau_R$) near the ballistic critical point ($1 - \tau \ll 1$), the differential capacitance reads (equations (49) and (52) in ref. 7; a factor $e\Delta/(2E_C)$ was applied to match the definition $C_{\text{diff}} \equiv \partial \langle Q \rangle / \partial V_g$):

$$C_{\text{diff}}^{1 - \tau \equiv 1 - \tau_L = 1 - \tau_R \ll 1} = \frac{4\gamma e}{\pi \Delta} \ln\left[(1 - \tau) \sin^2\left(\frac{\pi \delta V_g}{\Delta}\right) + \frac{k_B T}{E_C} \right] \times (1 - \tau) \cos\left(\frac{2\pi \delta V_g}{\Delta}\right) + \frac{e}{\Delta}$$

When approaching the ballistic critical point ($\tau \rightarrow 1$), the visibility in the oscillations of the differential capacitance asymptotically vanishes as $1 - \tau$, as in equation (7) with $\tau_L = \tau_R$.

Predictions in the presence of strong thermal fluctuations. In the presence of strong thermal fluctuations, $k_B T \gg E_C/\pi^2$, charge discreteness leads to periodic oscillations of the observables (for example, conductance and differential capacitance) while sweeping a capacitively coupled gate voltage. Quantum fluctuations decrease the oscillations, which are further attenuated by thermal fluctuations for increasing temperature, until the amplitude becomes exponentially small for $k_B T \gg E_C/\pi^2$. The exponential temperature dependence in $k_B T/E_C$ is quite robust, applying to thermodynamic^{1,8,29} and transport (Methods) properties. It can be demonstrated in the limits of both small and large transmission probabilities of the conduction channels comprising the junctions, and for various models of the metallic island. The presence of thermal fluctuations not only preserves the quantum $\sqrt{1 - \tau}$ suppression of the oscillations, but it is expected from the results of ref. 8 that the square-root scaling of the differential capacitance extends with increasing temperature, up to the full range of $\tau_{L,R} \in [0, 1]$. The relative oscillations in the differential capacitance and in the conductance characterize the degree of charge quantization equally well, both following the same $\exp(-\pi^2 k_B T/E_C) \sqrt{(1 - \tau_L)(1 - \tau_R)}$ behaviour. Further information regarding the predictions and theoretical methods in the presence of strong thermal fluctuations are provided in the following four sections.

Differential capacitance in the tunnel limit with strong thermal fluctuations. This regime corresponds to $k_B T \gg E_C/\pi^2$ and $\tau_{L,R} \ll 1$. To start with, we evaluate the oscillatory part of the free energy of the island in the limit $\tau_{L,R} \ll 1$, where the suppression of charge quantization is entirely due to thermal fluctuations. Considering high temperatures, it is convenient to transform the partition function of the isolated island,

$$Z = \sum_{n=-\infty}^{\infty} \exp\left[-\frac{E_n(\mathcal{N})}{k_B T}\right], \quad E_n(\mathcal{N}) = E_C(n - \mathcal{N})^2$$

using the Poisson summation formula; the result is

$$Z = \sqrt{\frac{\pi k_B T}{E_C}} \sum_{k=-\infty}^{\infty} \exp(-2\pi i k \mathcal{N}) \exp\left(-\frac{\pi^2 k_B T}{E_C} k^2\right) \quad (8)$$

Here $\mathcal{N} \equiv V_g/\Delta$ (with Δ the period in gate voltage V_g) is the charge induced by the gate voltage in units of e , and the summations are performed over integers n and k . The $k = 0$ and $k = \pm 1$ terms in the sum in equation (8) yield, respectively, the leading \mathcal{N} -independent and \mathcal{N} -dependent contributions F_0 and $\delta F(\mathcal{N})$ to the free energy $F = -k_B T \ln(Z)$ at $k_B T \gg E_C/\pi^2$. The resulting oscillatory part of the differential capacitance,

$$C_{\text{diff}}^{\tau_{L,R} \ll 1} \equiv \frac{e}{\Delta} \left(1 - \frac{1}{2E_C} \partial_{\mathcal{N}}^2 F \right) = \frac{e}{\Delta} - 4 \frac{e}{\Delta} \frac{\pi^2 k_B T}{E_C} \exp\left(-\frac{\pi^2 k_B T}{E_C}\right) \cos(2\pi \mathcal{N}) \quad (9)$$

is exponentially suppressed at high temperatures.

Differential capacitance in the near-ballistic regime with strong thermal fluctuations. This regime corresponds to $k_B T \gg E_C/\pi^2$ and $1 - \tau_R \ll 1$. A similar suppression of oscillations of the thermodynamic characteristics can also be demonstrated in the case of high-transmission junctions, where both thermal and quantum fluctuations contribute to the reduction of charge quantization. For definiteness, we consider here a single-junction case ($\tau_L = 0$) with $1 - \tau_R \ll 1$. Evaluation of $C_{\text{diff}}^{\tau_L=0, 1 - \tau_R \ll 1}$ can be performed using the bosonization scheme developed in ref. 7. In that formalism, the \mathcal{N} -dependent part of the differential capacitance reads

$$\delta C_{\text{diff}}^{\tau_L=0, 1 - \tau_R \ll 1} = -\frac{e}{\Delta} \frac{2\pi^2}{E_C} D \sqrt{1 - \tau_R} \langle \cos[2\pi \mathcal{N} - \varphi(0)] \rangle$$

where the bosonic quantum field $\varphi(0) = 2\pi \hat{Q}/e$ corresponds to the charge \hat{Q} passed through the junction ($x=0$), and D is the energy bandwidth appearing in the definition of boson variables. Averaging $\langle \dots \rangle$ is performed over the fluctuations of the field $\varphi(x)$. The Hamiltonian describing these fluctuations consists of two parts⁷, representing the energy of particle-hole excitations and the charging energy, respectively. The former part depends on $(\nabla \varphi)^2$ and the latter part has the form $E_C [\varphi(0)/(2\pi)]^2$. Replacement of the ground-state averaging⁷ with an average over the Gibbs distribution of fluctuations, which is proportional to $\exp\{-[E_C/(k_B T)] [\varphi(0)/(2\pi)]^2\}$, results in the renormalization of the bandwidth D to a physically meaningful value of approximately $k_B T$ and in exponential suppression of the oscillations at $k_B T \gg E_C/\pi^2$:

$$\delta C_{\text{diff}}^{\tau_{\text{L}}=0, 1-\tau_{\text{R}} \ll 1} \approx -4 \frac{e}{\Delta} \frac{\pi^2 k_{\text{B}} T}{E_{\text{C}}} \exp\left(-\frac{\pi^2 k_{\text{B}} T}{E_{\text{C}}}\right) \sqrt{1-\tau_{\text{R}}} \cos(2\pi \mathcal{N}) \quad (10)$$

Because it follows from ref. 8, equation (10) is applicable in the full range of τ_{R} for $k_{\text{B}} T \gg E_{\text{C}}/\pi^2$ (the numerical coefficient in equation (10) was established with the help of ref. 8). The identical exponential suppression for an almost-isolated island (equation (9)) is therefore simply equation (10) in the limit $\tau_{\text{R}} \ll 1$. In addition, quantum fluctuations contribute to the same suppression factor $\sqrt{1-\tau_{\text{R}}}$ derived at $1-\tau_{\text{R}} \ll 1$ in the quantum regime $k_{\text{B}} T \ll E_{\text{C}}$ (equation (5)). Furthermore, equation (10) derived for $k_{\text{B}} T \gg E_{\text{C}}$ matches the $T=0$ result of ref. 7 at $k_{\text{B}} T \approx E_{\text{C}}$; however, given the large numerical factor π^2 in the exponent of equation (10), there may be a broad crossover temperature region between the two limits.

Conductance in the tunnel limit with strong thermal fluctuations. This regime corresponds to $k_{\text{B}} T \gg E_{\text{C}}/\pi^2$ and $\tau_{\text{L,R}} \ll 1$. Turning now to conductance oscillations, we again start from the simpler case of low-transmission barriers ($\tau_{\text{L,R}} \ll 1$). In that limit, the rate equation for current carried by spin-polarized electrons yields³⁸:

$$G_{\text{SET}}^{\tau_{\text{L,R}} \ll 1}(\mathcal{N}, T) = \frac{e^2}{h} \frac{\tau_{\text{L}} \tau_{\text{R}}}{\tau_{\text{L}} + \tau_{\text{R}}} \sum_{n=-\infty}^{\infty} \left\{ \frac{\exp[-E_n(\mathcal{N})/(k_{\text{B}} T)]}{Z(\mathcal{N}, T)} \times f\left(\frac{E_n(\mathcal{N}) - E_{n-1}(\mathcal{N})}{k_{\text{B}} T}\right) \right\}$$

where $f(x) = x/(1 - e^{-x})$. Application of the Poisson summation formula to the above equation is tedious, but straightforward. The result is an expression for $G_{\text{SET}}^{\tau_{\text{L,R}} \ll 1}$ that involves a sum of harmonics proportional to $\cos(2\pi k \mathcal{N})$, similar to equation (8). The largest term,

$$G_{\infty} = \frac{e^2}{h} \left(\frac{1}{\tau_{\text{L}}} + \frac{1}{\tau_{\text{R}}} \right)^{-1} \quad (11)$$

does not oscillate and is simply the conductance of two resistors connected in series. The leading oscillatory term,

$$\delta G_{\text{SET}}^{\tau_{\text{L,R}} \ll 1}(\mathcal{N}, T) = -G_{\infty} \frac{\pi^2 k_{\text{B}} T}{E_{\text{C}}} \exp\left(-\frac{\pi^2 k_{\text{B}} T}{E_{\text{C}}}\right) \cos(2\pi \mathcal{N})$$

exhibits the same exponential suppression as the differential capacitance (equation (9)).

Conductance in the near-ballistic regime with strong thermal fluctuations. This regime corresponds to $k_{\text{B}} T \gg E_{\text{C}}/\pi^2$ and $1 - \tau \ll 1$. Regarding the conductance across a metallic island with high-transmission contacts, we (A) present a formalism that is somewhat different from ref. 25, details of which will be published separately (E.I., I.P.L. and E.V.S., manuscript in preparation) and (B) further establish the predictions by extending the formalism of ref. 25 to high temperatures.

(A) In the first approach, we start from the chiral edge excitations of the integer quantum Hall regime, in close correspondence with the experimental configuration. Although we are interested in the high-temperature limit, all the energy scales in the experiment remain much smaller than the quantum Hall energy gap. At such low energies, the quantum Hall edge states may be described by the effective theory^{39–41}. According to this theory, edge excitations can be viewed as bosonic edge magneto-plasmons. The corresponding one-dimensional charge density waves $\rho_{s\alpha}(x)$ ($s \in \{\text{L}, \text{R}\}$, $\alpha \in \{1, 2\}$; see Extended Data Fig. 4 for notations) verify the canonical commutation relations $[\rho_{s\alpha}(x), \rho_{s'\beta}(y)] = (-1)^{\alpha} 2\pi i e^2 \delta_{ss'} \delta_{\alpha\beta} \delta'(x - y)$, where the sign accounts for the propagation direction of the chiral edge states, $\delta_{s's'}$ and $\delta_{\alpha\beta}$ are Kronecker delta functions, and $\delta'(x - y)$ is the derivative of the Dirac delta function.

The Hamiltonian of the experimental set-up contains three terms: $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{\text{int}} + \mathcal{H}_{\text{T}}$. The first term describes the dynamics of the bare edge states:

$$\mathcal{H}_0 = \frac{\hbar v_{\text{F}}}{2e^2} \sum_{s\alpha} \int dx \rho_{s\alpha}^2(x)$$

where v_{F} is the Fermi velocity of the quantum Hall edge states. The second term describes Coulomb interactions at the metallic island:

$$\mathcal{H}_{\text{int}} = E_{\text{C}}(\hat{Q}/e - \mathcal{N})^2 \quad (12)$$

$$\hat{Q} = \frac{e}{2\pi} [\varphi_{\text{L}}(0) - \varphi_{\text{R}}(0)] = \sum_{\alpha} \left[\int_0^{\infty} dx \rho_{\text{L}\alpha}(x) + \int_{-\infty}^0 dx \rho_{\text{R}\alpha}(x) \right] \quad (13)$$

The first equality in equation (13) defines the Bose field operators that are also used in the derivation of equation (10), but here for the case of two contacts. The last term in the Hamiltonian describes the backscattering of electrons at the two QPCs:

$$\mathcal{H}_{\text{T}} = A_{\text{L}} + A_{\text{R}} + \text{h.c.}$$

$$A_s = \gamma_s \psi_{s1}^{\dagger}(0) \psi_{s2}(0) \quad (14)$$

$$\psi_{s\alpha}(0) = \sqrt{\frac{D}{\hbar v_{\text{F}}}} \exp\left[\frac{2\pi i}{e} \int_{-\infty}^0 dx \rho_{s\alpha}(x)\right] \quad (15)$$

where the backscattering amplitudes $\gamma_{\text{L,R}}$ depend on the ‘intrinsic’ transmission probabilities $\tau_{\text{L,R}}$ (in the near ballistic regime, $1 - \tau_{\text{L(R)}} \approx |\gamma_{\text{L(R)}}|^2/(\hbar v_{\text{F}})^2$).

We set the distance between the metallic island and the QPCs, which is much shorter than the wavelength of excitations in the experiment, to zero. We stress that exactly the same Hamiltonian arises in the absence of the quantum Hall effect, when applying the bosonization procedure to a metallic island connected to reservoirs through spin-polarized electron channels (as in refs 7 and 25). Consequently, the predictions below apply beyond the quantum Hall configuration used here as a starting point.

Focusing on the near ballistic regime $1 - \tau_{\text{L,R}} \ll 1$, we apply the scattering theory approach developed in refs 42 and 43. The average $\langle I \rangle \equiv \text{tr}(\rho I)$ of the current operator $I = v_{\text{F}}[\rho_{\text{R}1}(0) - \rho_{\text{R}2}(0)]$ is evaluated perturbatively in backscattering amplitudes (equation (14)). With this aim, we express the density matrix $\rho = U \rho_0 U^{\dagger}$ in terms of its equilibrium value $\rho_0 \propto \exp[-(\mathcal{H}_0 + \mathcal{H}_{\text{int}})/(k_{\text{B}} T)]$, and expand the evolution operator $U = \hat{T} \exp[-2\pi i/\hbar \int dt \mathcal{H}_{\text{T}}(t)]$ in powers of γ_s , where $\hat{T} \exp$ indicates the time-ordered exponential. This results in the two leading terms:

$$\langle I \rangle = \langle I \rangle_0 + \frac{1}{\hbar^2} \left\langle \iint dt' dt'' \{ \mathcal{H}_{\text{T}}(t'), [\mathcal{H}_{\text{T}}(t''), I] \} \right\rangle_0 \quad (16)$$

where the average is taken with respect to the equilibrium density matrix ρ_0 .

The Hamiltonian $\mathcal{H}_0 + \mathcal{H}_{\text{int}}$ is quadratic in plasmon operators. Consequently, the corresponding dynamics can be accounted for exactly within the scattering theory approach for bosons^{42,43}. For instance, the scattering matrix for the interaction Hamiltonian \mathcal{H}_{int} (ignoring the backscattering Hamiltonian \mathcal{H}_{T}), which relates the currents in the incoming (L1, R2, L2, R1) and outgoing (L2, R1, L1, R2) channels at the frequency $\omega/(2\pi)$, reads:

$$S(\omega) = \frac{1}{2} \begin{pmatrix} z & z & 2-z & -z \\ z & z & -z & 2-z \\ 2-z & -z & z & z \\ -z & 2-z & z & z \end{pmatrix} \quad (17)$$

where $z = 1/[i\hbar\omega/(4E_{\text{C}}) + 1]$. Taking the limit $\omega \rightarrow 0$ (ref. 43), we determine the first term in equation (16): $\langle I \rangle_0 = e^2 V_{\text{dc}}/(2\hbar)$. The bare conductance is thus half the conductance quantum. In the limit of small d.c. bias V_{dc} , the second term can be rewritten as

$$\delta \langle I \rangle = \frac{e}{\hbar^2} \int dt \langle [A_{\text{L}}^{\dagger}(t) + A_{\text{R}}^{\dagger}(t), A_{\text{L}}(0) + A_{\text{R}}(0)] \rangle_0$$

This term contains the coherent contribution

$$\delta \langle I \rangle_{\text{osc}} = \frac{e}{\hbar^2} \text{Re} \gamma_{\text{L}}^* \gamma_{\text{R}} \int dt \langle \psi_{\text{L}2}^{\dagger}(0, t) \psi_{\text{L}1}(0, t) \psi_{\text{R}1}^{\dagger}(0, 0) \psi_{\text{R}2}(0, 0) \rangle_0 \quad (18)$$

which oscillates as a function of the induced charge $e\mathcal{N}$.

In general, one can use the scattering matrix equation (17) to evaluate the average in equation (18), which leads to a complex expression (E.I., I.P.L. and E.V.S., manuscript in preparation). However, the leading high-temperature asymptotics can be found using exactly the same argument as for the case of the differential capacitance considered above. Specifically, according to equation (15), the particular value of the charge Q in the island leads to the phase shift of $\exp[2\pi i(Q/e - \mathcal{N})]$ in the correlation function in equation (18). Therefore, by averaging the correlation function over instant fluctuations of this charge, which are distributed with the equilibrium Gibbs weights (proportional to $\exp[-(Q/e)^2 E_{\text{C}}/(k_{\text{B}} T)]$), we determine the high-temperature behaviour of the oscillating part of the current:

$$\begin{aligned} \delta \langle I \rangle_{\text{osc}} &\propto \frac{e^2 V_{\text{dc}}}{h} \sqrt{\frac{E_{\text{C}}}{k_{\text{B}} T}} \sqrt{(1-\tau_{\text{L}})(1-\tau_{\text{R}})} \\ &\times \int dQ \exp\left(-\frac{Q^2 E_{\text{C}}}{e^2 k_{\text{B}} T}\right) \cos[2\pi(\mathcal{N} - Q/e)] \\ &\propto \frac{e^2 V_{\text{dc}}}{h} \exp\left(-\frac{\pi^2 k_{\text{B}} T}{E_{\text{C}}}\right) \sqrt{(1-\tau_{\text{L}})(1-\tau_{\text{R}})} \cos(2\pi \mathcal{N}) \end{aligned}$$

The validity of this simplified approach is confirmed by calculations (E.I., I.P.L. and E.V.S., manuscript in preparation).

(B) An alternative route of calculation amounts to re-working equation (A5) of ref. 25 for the case $1 - \tau_{L,R} \ll 1$ or equation (A27) for the asymmetric case $\tau_L \ll 1$, $1 - \tau_R \ll 1$. In either case, the largest term in the limit $k_B T \gg E_C/\pi^2$ is, unsurprisingly, \mathcal{N} -independent. Like equation (11), it represents the conductance of two junctions connected in series: $G_\infty \approx e^2/(2h)$ in the case of $1 - \tau_{L,R} \ll 1$ and $G_\infty \approx (e^2/h)\tau_L$ in the case $\tau_L \ll 1$, $1 - \tau_R \ll 1$. The leading oscillatory term in the former case is

$$\delta C_{\text{SET}}^{1-\tau_{L,R} \ll 1}(\mathcal{N}, T) \approx \frac{e^2}{h} \exp\left(-\frac{\pi^2 k_B T}{E_C}\right) \sqrt{(1-\tau_L)(1-\tau_R)} \cos(2\pi\mathcal{N})$$

In the asymmetric case, the factor $\sqrt{(1-\tau_L)(1-\tau_R)}$ in the above expression is replaced by $\tau_L \sqrt{1-\tau_R}$. The visibility of conductance oscillations now reads:

$$\Delta Q \approx \exp\left(-\frac{\pi^2 k_B T}{E_C}\right) \sqrt{(1-\tau_L)(1-\tau_R)}$$

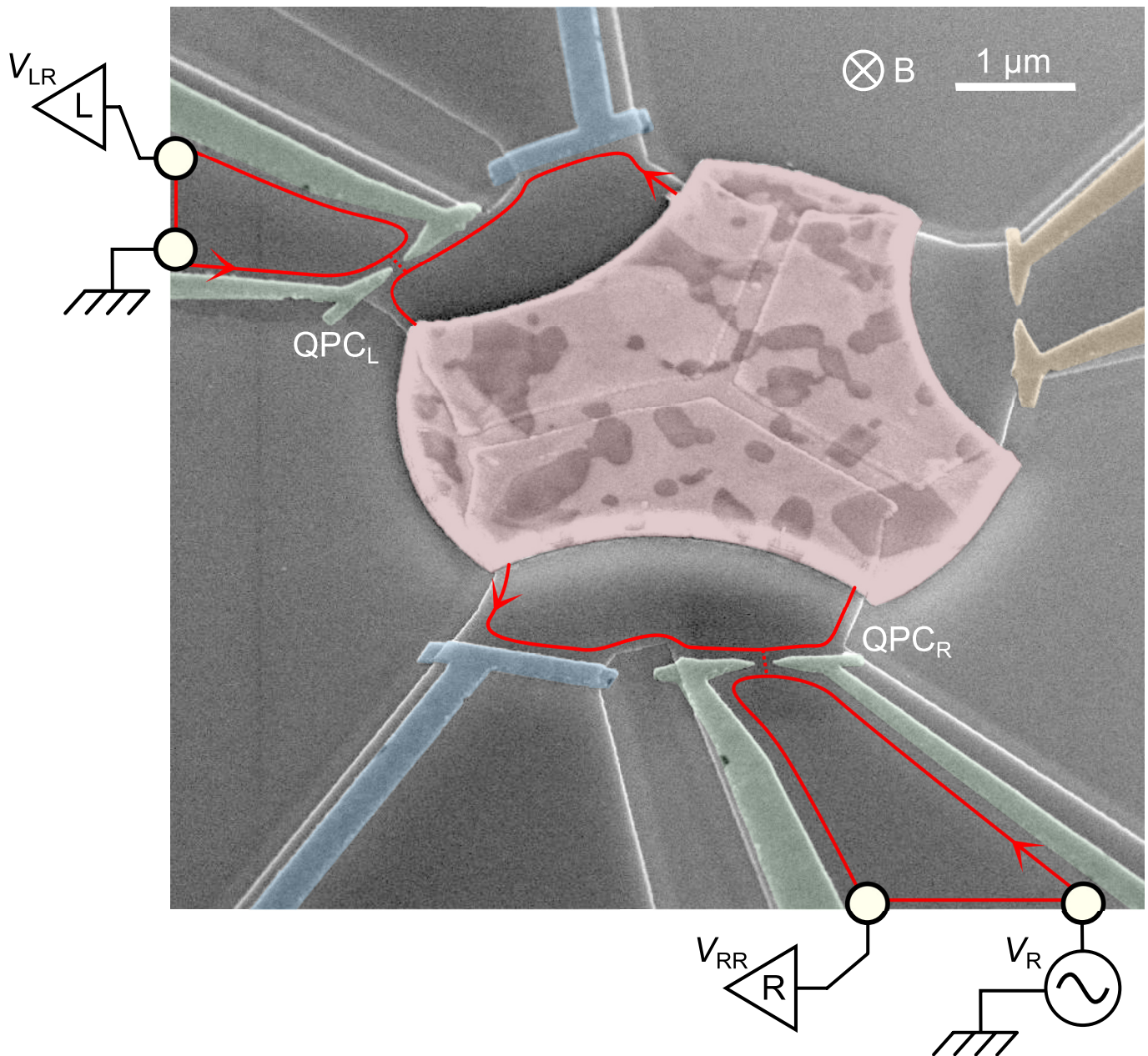
This form correctly extrapolates between the symmetric and asymmetric cases.

Conductance at $T \approx 17$ mK versus quantum regime predictions. Although the visibility ΔQ of the oscillations in the SET conductance best reflects the degree of charge quantization, we can also confront experiment and theory directly at the underlying conductance-sweeps level. In Extended Data Fig. 3, we compare $G_{\text{SET}}(\delta V_g)$ measurements (symbols) and predictions near the ballistic critical point ($1 - \tau_R \approx 0.02$ and 0.004) with QPCs in both the tunnel ($\tau_L = 0.075$) and almost perfectly transmitted ($1 - \tau_L \approx 0.02$) regimes. Solid lines are calculated with the electronic temperature $T = 17$ mK, using equation (3) for the top two panels (asymmetric regime, $\tau_L = 0.075$) and equation (6) for the bottom two panels (near ballistic regime, $\tau_L = 0.983$). The grey areas correspond to the experimental uncertainty of ± 4 mK. The demonstrated agreement validates the full prediction for the renormalized SET conductance.

Charge quantization based on conductance or transmission probability values. Theory predicts that as soon as one conduction channel connected to the metallic island is ballistic, the charge in the island is completely unquantized. Here we show that charge quantization collapses systematically at the ballistic critical point $\tau_R = 1$, independent of the setting of the second channel ($\tau_L < 1$). We further demonstrate that the crucial ingredient is not the overall conductance, but the presence of a perfectly transmitted channel. For this purpose, we compare the two configurations displayed in Extended Data Fig. 5a, b. In both configurations, QPC_L is tuned to the same standard setting corresponding to a single conduction channel of ‘intrinsic’ transmission probability $\tau_L = 0.24$. In both configurations, QPC_R is set to the same

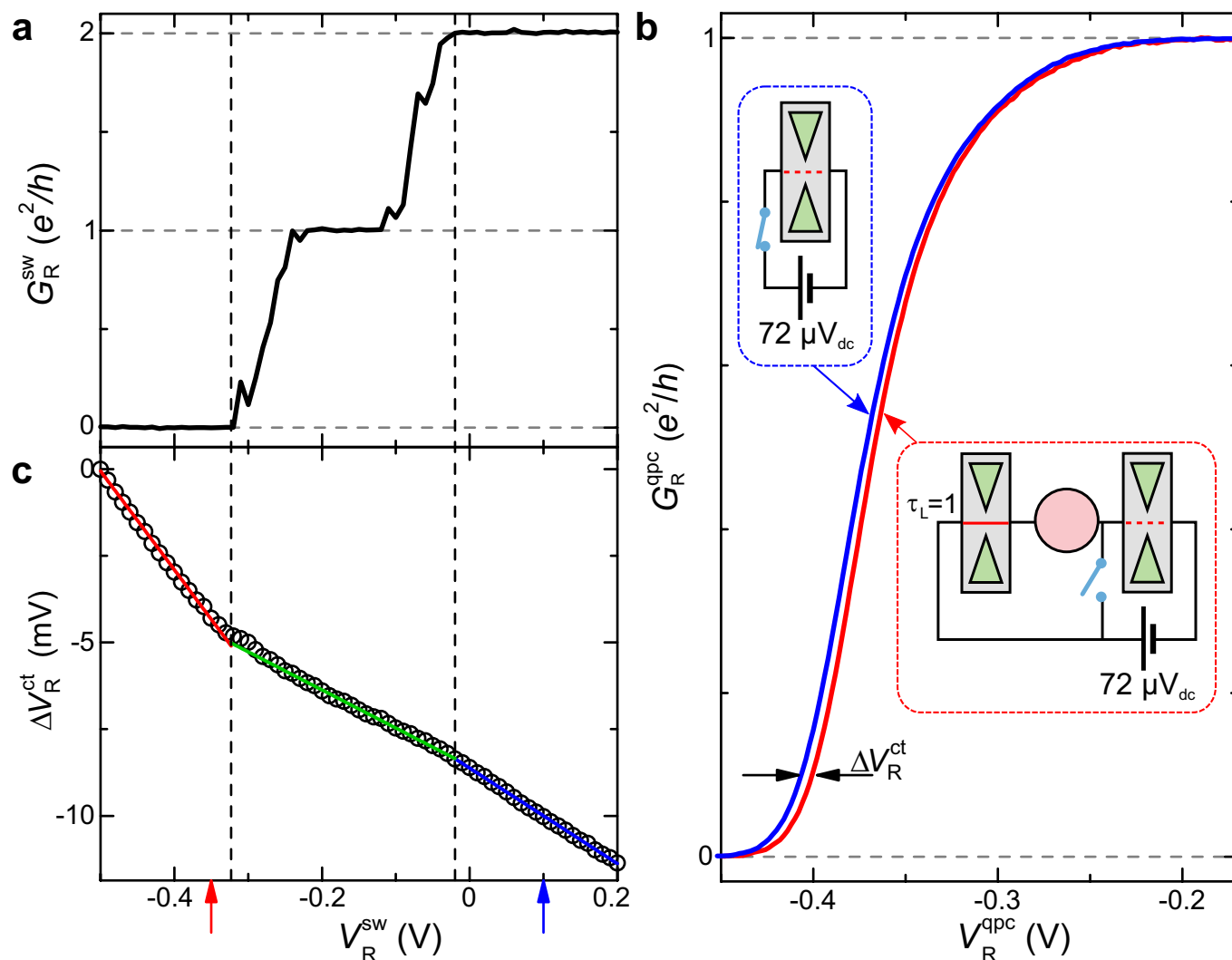
overall intrinsic conductance $G_R^{\text{qpc}} \equiv \tau_R e^2/h = 1.5e^2/h$. However, in the first configuration (Extended Data Fig. 5a) QPC_R decomposes into one ballistic channel and one channel of intrinsic transmission probability 0.5, whereas in the second configuration (Extended Data Fig. 5b) it decomposes into two non-ballistic channels of intrinsic transmission probabilities 0.7 and 0.8. (In practice, the QPC_R of the second configuration is realized using two different physical QPCs biased at the same voltage.) As shown Extended Data Fig. 5c, the SET conductance displays strong oscillations in the second configuration, signalling charge quantization in the absence of a ballistic channel. By contrast, the SET conductance in the first configuration does not depend on gate voltage, signalling a completely unquantized island charge in the presence of one ballistic channel.

31. Göktaş, O., Weber, J., Weis, J. & von Klitzing, K. Alloyed ohmic contacts to two-dimensional electron system in AlGaAs/GaAs heterostructures down to submicron length scale. *Physica E* **40**, 1579–1581 (2008).
32. Iftikhar, Z. *et al.* Two-channel Kondo effect and renormalization flow with macroscopic quantum charge states. *Nature* **526**, 233–236 (2015).
33. Jezouin, S. *et al.* Quantum limit of heat flow across a single electronic channel. *Science* **342**, 601–604 (2013).
34. Spietz, L., Lehnert, K. W., Siddiqi, I. & Schoelkopf, R. J. Primary electronic thermometry using the shot noise of a tunnel junction. *Science* **300**, 1929–1932 (2003).
35. Panyukov, S. V. & Zaikin, A. D. Coulomb blockade and nonperturbative ground-state properties of ultrasmall tunnel junctions. *Phys. Rev. Lett.* **67**, 3168–3171 (1991).
36. Wang, X. & Grabert, H. Coulomb charging at large conduction. *Phys. Rev. B* **53**, 12621–12624 (1996).
37. Lukyanov, S. L. Notes on parafermionic QFTs with boundary interaction. *Nucl. Phys. B* **784**, 151–201 (2007).
38. Glazman, L. I. & Shekhter, R. I. Coulomb oscillations of the conductance in a laterally confined heterostructure. *J. Phys. Condens. Matter* **1**, 5811–5815 (1989).
39. Halperin, B. I. Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential. *Phys. Rev. B* **25**, 2185–2190 (1982).
40. Wen, X. G. Chiral Luttinger liquid and the edge excitations in the fractional quantum Hall states. *Phys. Rev. B* **41**, 12838–12844 (1990).
41. Fröhlich, J. & Zee, A. Large scale physics of the quantum Hall fluid. *Nucl. Phys. B* **364**, 517–540 (1991).
42. Slobodeniuk, A., Levkivskyi, I. & Sukhorukov, E. Equilibration of quantum Hall edge states by an Ohmic contact. *Phys. Rev. B* **88**, 165307 (2013).
43. Sukhorukov, E. Scattering theory approach to bosonization of non-equilibrium mesoscopic systems. *Physica E* **77**, 191–198 (2016).



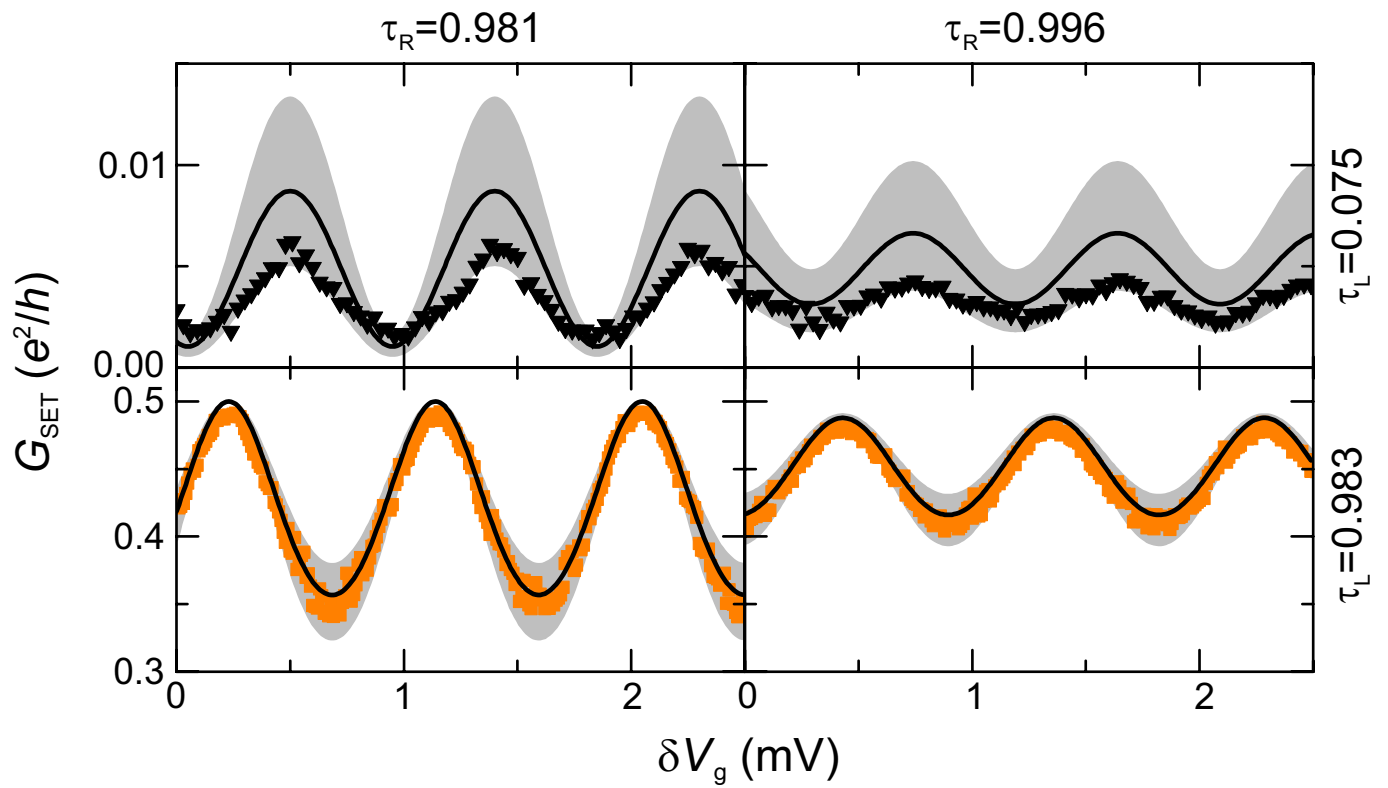
Extended Data Figure 1 | Measurement schematic. The signal V_{LR} (V_{RR}) is the voltage measured with amplification chain L (R) in response to the injected voltage V_R . The trenches etched in the 2DEG, which can be seen in the form of a 'Y' through the metallic island, ensure that the only way

from one QPC to the other is across the metallic island. The experiment is performed in the quantum Hall regime at filling factor $\nu=2$, where the current propagates along the edges in the direction indicated by arrows.

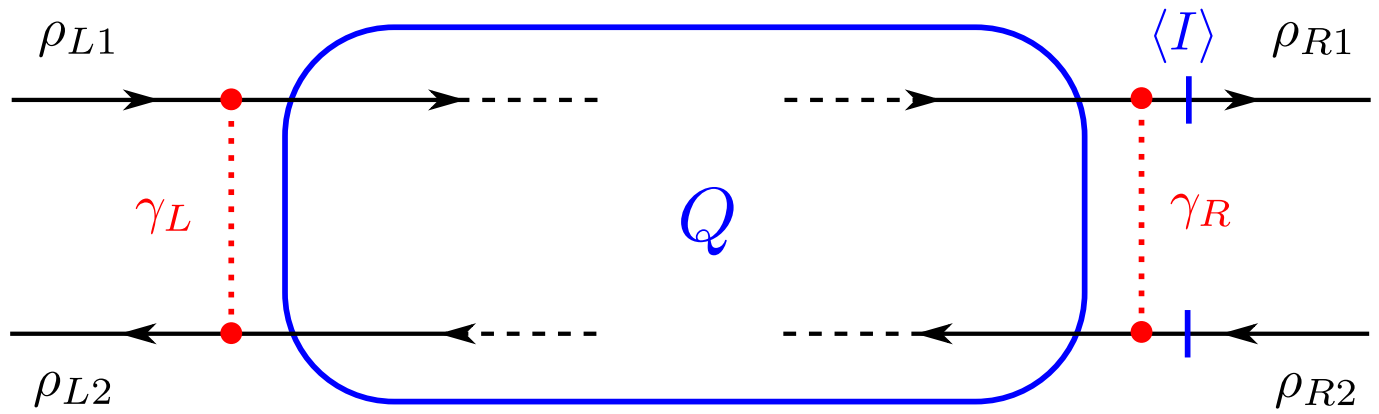


Extended Data Figure 2 | Crosstalk compensation. **a**, (Intrinsic) conductance G_R^{sw} across the characterization gate adjacent to QPC_R versus gate voltage V_R^{sw} . In the experiment, the left and right switches are independently set to the open and closed positions with $V_{R,L}^{sw} = -0.35$ V and $V_{R,L}^{sw} = 0.1$ V, respectively (vertical arrows in **c**). **b**, QPC_R differential conductance in the presence of a d.c. bias of $72 \mu V$ ($72 \mu V_{dc}$) versus QPC gate voltage V_R^{qpc} . The red and blue lines are measured with the adjacent switch in the open and closed positions, respectively (see inset schematics). The voltage drop across QPC_R is smaller with the switch

open, owing to the added series resistance. Although this does not result in a large error, because G_R^{qpc} depends weakly on voltage bias, this effect is minimized by extracting the crosstalk compensation ΔV_R^{ct} at low $G_R^{qpc} \lesssim 0.1 e^2/h$. **c**, Symbols represent the crosstalk compensation ΔV_R^{ct} , with respect to the gate voltage $V_R^{sw} = -0.5$ V, versus V_R^{sw} . Lines are linear fits of the crosstalk compensation at $G_R^{sw} = 0$ (red, -2.8% relative compensation), $0 < G_R^{sw} < 2 e^2/h$ (green, -1.1% relative compensation) and $G_R^{sw} = 2 e^2/h$ (blue, -1.4% relative compensation).

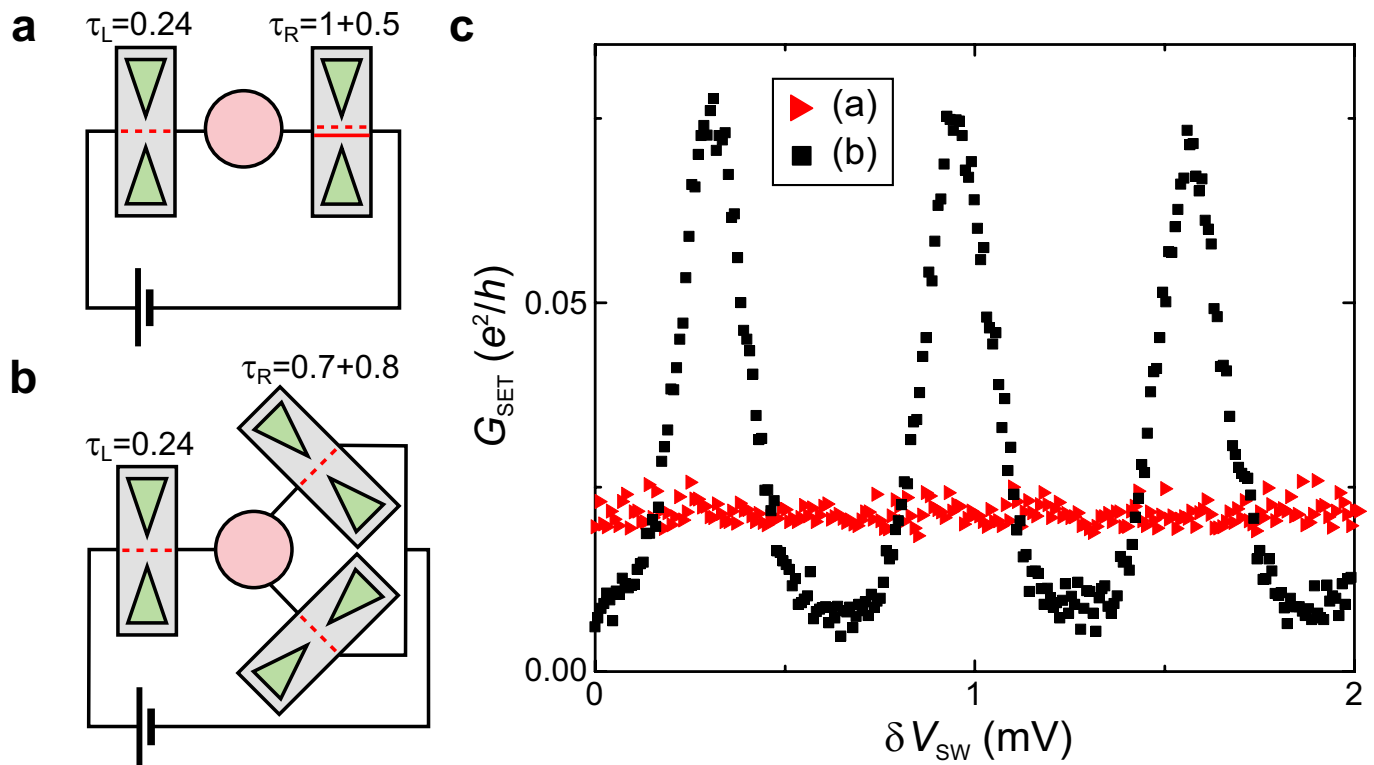


Extended Data Figure 3 | Conductance measurements versus quantitative predictions. Direct $G_{\text{SET}}(\delta V_g)$ comparison at $T \approx 17$ mK between data (symbols) and predictions (solid lines, grey areas correspond to the temperature uncertainty of ± 4 mK) in the two limits addressed by theory (equation (3) for $\tau_L \approx 0$ (top panels), equation (6) for $\tau_L \approx 1$ (bottom panels)).



Extended Data Figure 4 | Theoretical description of the experimental set-up in formalism (A) for strong thermal fluctuations. We consider the regime of the quantum Hall effect, where only one spinless edge mode contributes to the transport. The corresponding edge states are described by four charge density operators, labelled by $s \in \{L, R\}$ and $\alpha \in \{1, 2\}$. These states are mixed (backscattered) at the two QPCs (red dashed lines)

with amplitudes γ_L and γ_R (equations (14) and (15)). The edge densities enter into the interaction Hamiltonian (equation (12)) through the total charge \hat{Q} of the metallic island (equation (13)). The average current $\langle I \rangle$ is calculated through a cross-section immediately to the right of QPC_R (vertical blue lines).



Extended Data Figure 5 | Charge quantization based on conductance versus transmission probability values. **a, b**, Schematics of the configurations, both with the same QPC_L setting $\tau_L=0.24$. In the configuration shown in **a**, QPC_R is set to an ‘intrinsic’ conductance $G_R^{\text{qpc}} \equiv \tau_R e^2/h = 1.5e^2/h$, which decomposes into one ballistic channel and one channel of intrinsic transmission probability 0.5. In the configuration shown in **b**, QPC_R is set

to the same intrinsic conductance $G_R^{\text{qpc}} = 1.5e^2/h$, which now decomposes into two non-ballistic channels of intrinsic transmission probabilities 0.7 and 0.8. **c**, Sweeps of the device conductance are plotted versus gate voltage for the two configurations (**a**, red triangles; **b**, black squares). Conductance oscillations are visible only in the configuration shown in **b**, in the absence of a ballistic channel connected to the island.

Demonstration of a small programmable quantum computer with atomic qubits

S. Debnath¹, N. M. Linke¹, C. Figgatt¹, K. A. Landsman¹, K. Wright¹ & C. Monroe^{1,2,3}

Quantum computers can solve certain problems more efficiently than any possible conventional computer. Small quantum algorithms have been demonstrated on multiple quantum computing platforms, many specifically tailored in hardware to implement a particular algorithm or execute a limited number of computational paths^{1–10}. Here we demonstrate a five-qubit trapped-ion quantum computer that can be programmed in software to implement arbitrary quantum algorithms by executing any sequence of universal quantum logic gates. We compile algorithms into a fully connected set of gate operations that are native to the hardware and have a mean fidelity of 98 per cent. Reconfiguring these gate sequences provides the flexibility to implement a variety of algorithms without altering the hardware. As examples, we implement the Deutsch–Jozsa¹¹ and Bernstein–Vazirani¹² algorithms with average success rates of 95 and 90 per cent, respectively. We also perform a coherent quantum Fourier transform^{13,14} on five trapped-ion qubits for phase estimation and period finding with average fidelities of 62 and 84 per cent, respectively. This small quantum computer can be scaled to larger numbers of qubits within a single register, and can be further expanded by connecting several such modules through ion shuttling¹⁵ or photonic quantum channels¹⁶.

Implementing a scalable programmable quantum computing architecture requires high-fidelity initialization and detection at the individual qubit level and pristine control of interactions between qubits. Whereas most physical platforms have nearest-neighbour interactions only, a multi-qubit trapped-ion system features an intrinsic long-range interaction that is optically gated and connects any pair of qubits^{17,18}. Unlike solid-state implementations^{19,20}, the quantum circuitry is determined by external fields, and hence can be programmed and reconfigured without altering the structure of the qubits themselves. By optically resolving individual ions, we implement single-qubit rotations and arbitrary two-qubit gates by directly addressing pairs of ions without additional overhead such as moving information through local couplings^{19,21} or hiding qubit populations in additional auxiliary states¹⁰. Such native gates can then be used to construct modular logic gates that can be called in reconfigurable algorithm sequences. We observe a mean fidelity of 98% in these native operations without the use of spin echo or dynamical decoupling techniques^{10,22,23}. This bottom-up approach can be adapted for large-scale computation using micro-fabricated ion traps with integrated optics²⁴ and high optical access, and we expect that gate fidelities could exceed 99.9% with straightforward improvements to the classical control^{23,25}.

The programmable and reconfigurable nature of the trapped-ion quantum computer is illustrated by a hierarchy of operations from software to hardware, shown in Fig. 1a. At the top is a high-level user interface that specifies the desired algorithm, represented by a standard family of modular universal logic gates such as Hadamard (H), controlled-NOT (CNOT), and controlled-phase (CP) gates¹⁴. Next, a quantum compiler translates the universal gates into gates native to the hardware, which in our case are two-qubit Ising (XX) gates¹⁸ and

single-qubit rotation (R) gates¹⁴. Finally, these native gates are decomposed into laser pulses that are pre-calculated to effect the desired qubit operation through the Coulomb-coupled motion while disentangling the motion at the end of the gates²⁶.

At the hardware level, the processor consists of trapped ¹⁷¹Yb⁺ atomic ion qubits with information stored in the hyperfine ‘clock’ states $|0\rangle \equiv |F=0; m_F=0\rangle$ and $|1\rangle \equiv |F=1; m_F=0\rangle$ of the ²S_{1/2} electronic ground level with a qubit frequency splitting of $\nu_0 = 12.642821$ GHz (ref. 27). Here, F and m_F denote the quantum numbers associated with the total atomic angular momentum and its projection along the quantization axis defined by an applied magnetic field of 5.2 G. We measure a qubit coherence time in excess of 0.5 s, and with magnetic shielding we expect this to improve dramatically (ref. 28).

We confine the ions in a linear radio-frequency Paul trap, with radial and axial trap frequencies $\nu_x = 3.07$ MHz and $\nu_z = 0.27$ MHz, respectively. The ions are laser-cooled to near their radial motional ground state and form a linear crystal with a spacing of approximately 5 μ m for $n = 5$ ions. A computation is performed by first initializing all qubits to state $|0\rangle$ through optical pumping²⁷. This is followed by quantum gates, implemented by a series of coherent rotations using stimulated Raman transitions driven by a 355 nm mode-locked laser, where the beat-note between two counterpropagating Raman beams drives qubit and motional transitions²⁹. To achieve individual addressing, we split one of the Raman beams into a static array of beams, each of which is directed through an individual channel of a multi-channel acousto-optic modulator (AOM) and focused onto its respective ion, as shown in Fig. 1b. Finally, the qubit register is measured with high fidelity (see Methods) by driving the ²S_{1/2} \rightarrow ²P_{1/2} cycling transition near 369 nm and simultaneously collecting the resulting state-dependent fluorescence from each ion using high-resolution optics and a multi-channel photo-multiplier tube (PMT).

The lowest level of qubit control consists of native single- and two-qubit operations. We perform single-qubit rotations $R_\phi(\theta)$ by tuning the Raman beat-note to qubit resonance ν_0 . Here, the rotation angle θ and axis ϕ are determined by the duration and phase offset of the beat-note, which is programmed through radio-frequency signals on appropriate AOM channels. Two-qubit XX-gates are performed by invoking an effective spin–spin Ising interaction between qubits mediated by the collective modes of motion of the chain¹⁸. Here, we apply Raman beat-notes tuned close to $\nu_0 \pm \nu_x$ that coherently couple the spins to all X-modes of motion. A pulse shaping technique²⁶ disentangles the motion at the end of the gate, resulting in a two-qubit entangling rotation of any amount $XX(\chi_{ij})$. Here, the geometric phase χ_{ij} originates from the integrated Ising interaction^{18,26}, the sign $\alpha_{ij} = \pm 1$ of which arises from the Coulomb interaction between qubits i and j (Fig. 1b inset). We pre-calculate and optimize XX-gate pulse shapes off-line for all $\{i, j\}$ to achieve high fidelity while keeping the gates relatively fast (see Methods).

We use these native R- and XX-gates to construct standard logic gates, which can be called by a quantum algorithm. For instance, we

¹Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA. ²Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, Maryland 20742, USA. ³IonQ, Inc., College Park, Maryland 20742 USA.

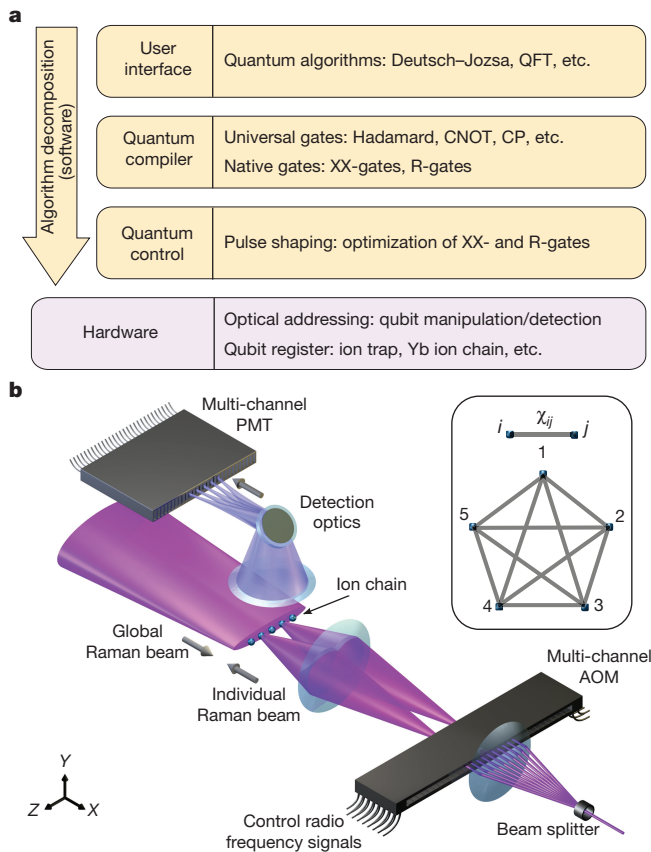


Figure 1 | Computation architecture. **a**, Hierarchy of operations from software to hardware. See main text for details. **b**, Hardware setup. A linear chain of trapped ion qubits along the Z axis is shown at the centre of the panel ('Ion chain'). An imaging objective ('Detection optics') collects ion fluorescence along the Y axis and maps each ion onto a multichannel photo-multiplier tube (PMT) for measurement of individual qubits. Counterpropagating Raman beams ('Global' and 'Individual') along the X axis perform qubit operations. A diffractive beam splitter creates an array of static Raman beams that are individually switched using a multi-channel acousto-optic modulator (AOM) driven by radio frequency ('Control radio-frequency signals') to perform qubit-selective gates. By modulating appropriate addressing beams, any single-qubit rotation or two-qubit Ising (XX) gate can be realized. For the two-qubit gates between qubits i and j , we can continuously tune the nonlinear gate angle χ_{ij} . This represents a system of qubits with fully connected and reconfigurable spin–spin Ising interactions (inset).

implement the single-qubit Hadamard gate as $H = R_x(-\pi)R_y(\pi/2)$ and the Z-rotation as $R_z(\theta) = R_y(-\pi/2)R_x(\theta)R_y(\pi/2)$. Two-qubit logic gates such as CP and CNOT are compiled to account for the signs of the CP rotation angle β and the Ising interaction α_{ij} , making them independent of $\{i, j\}$ and therefore modular (Fig. 2). At the highest level we program arbitrary sequences of such logic gates as required to implement any quantum algorithm.

We first implement the Deutsch–Jozsa algorithm¹¹, which determines whether a given function (the 'oracle') is constant or balanced. A function that has an n -bit input and a 1-bit output ($f: \{0, 1, 2, \dots, 2^n - 1\} \rightarrow \{0, 1\}$) is balanced when exactly half of the inputs result in the output 0 and the other half in the output 1, while a constant function assumes a single value irrespective of the input. In our setup we program 7 out of the 70 possible oracles of three-qubit balanced functions by using seven different sequences of CNOT gates between each of the three qubits in the control register $x = \{X_1X_2X_3\}$ and the function register X_4 (Fig. 3a). We program the two constant functions by setting X_4 to either 0 or 1. Executing the algorithm starts with preparing the control register in the superposition state $|x\rangle = \frac{1}{\sqrt{8}} \sum_{k=0}^7 |k\rangle$, followed

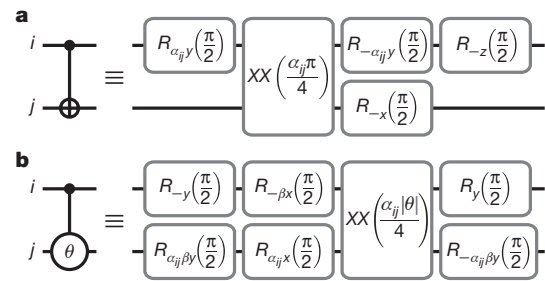


Figure 2 | Two-qubit modular gates. **a**, Decomposition of the controlled-NOT (CNOT) gate. The geometric phase χ_{ij} of the XX-gate is $\pm\pi/4$, and we define $\alpha_{ij} = \text{sgn}(\chi_{ij})$. **b**, Decomposition of the controlled-phase (CP) gate where $\beta = \text{sgn}(\theta)$ for the controlled phase θ . The geometric phase of the XX-gate is adjusted such that $\chi_{ij} = \alpha_{ij}|\theta|/4$. See main text for details.

by the function evaluation oracle. A CNOT is then performed between the function register X_4 and the ancilla qubit X_5 (initially set to $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$). All qubits are then rotated and measured (except for the ancilla) as shown in Fig. 3a. Finally, a measurement of x (conditioned upon $X_4 = 1$, occurring with 50% probability) determines if the function is constant or balanced. Measurement of the output $x = \{111\}$ indicates a constant function, while any other value indicates a balanced function (see Methods). The average success probability is 0.967(2) for constant and 0.932(3) for balanced functions (Fig. 3b), where the number in parentheses is the statistical uncertainty (1 s.d.).

The Bernstein–Vazirani algorithm is a variant of the Deutsch–Jozsa algorithm where the oracle function is an inner product of two n -bit strings: $f_c(x) = c \cdot x$. Here, the aim is to determine the vector $c = \{c_1c_2\dots c_n\}$ in a single trial¹². We program all 16 instances of the four-bit oracle that evaluate the function $f_c(x) \oplus X_5$. This is achieved by applying a particular pattern of CNOT gates, determined by c , between $x = \{X_1X_2X_3X_4\}$ and $X_5 = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ (Fig. 3c). For example, if $c = \{0101\}$ then CNOT gates are applied between X_2, X_5 and X_4, X_5 . We start with a superposition state $|x\rangle = \frac{1}{\sqrt{16}} \sum_{k=0}^{15} |k\rangle$, followed by the oracle. Finally, applying a global $R_y(\pi/2)$ rotation produces the output state \bar{c} , which is the inverse of c . In the experiment, a single-shot measurement of the correct outcome \bar{c} is obtained with a probability of 0.903(2) (Fig. 3d), averaged over all possible oracle states.

Exponential speed-up of many quantum algorithms arises from the fact that parallel function evaluation is performed on a superposition of all classical input states of an n -bit string. These evaluation paths are then interfered using a quantum Fourier transform (QFT) to produce the desired solution¹⁴. One such example is the order-finding protocol in Shor's quantum factorization algorithm¹³. Another application is solving the eigenvalue problem $A|\phi\rangle = e^{i\phi}|\phi\rangle$, where the phase ϕ can be estimated to n -bit precision using an n -bit QFT¹⁴. These algorithms have been implemented in experiments using a semi-classical version of the QFT that consists of single-qubit rotations based on classical feed-forward and qubit recycling which reduces the required register size^{5,10,30}. The coherent QFT, on the other hand, is reversible and can be concatenated within an algorithm sequence.

Here, we construct a coherent QFT on five qubits using all 10 modular CP gates and involving a total of 80 single- and two-qubit native gates. This circuit fully exploits the high connectivity of a trapped ion system and illustrates how it can be scaled to larger modules (Fig. 4a). We apply the QFT in a period-finding protocol where we first prepare an input superposition state $\sum_{k=0}^{31} C_k|k\rangle$ such that the coefficients $\{C_k\}$ exhibit a periodic amplitude or phase modulation (see Methods), which is followed by the QFT operation. The modulation periodicity then appears in the output state populations (Fig. 4b).

We further examine the performance of the QFT in a phase estimation protocol where the eigenvalue ϕ is estimated to 5-bit precision. In this case the input state is prepared in the form $\frac{1}{\sqrt{32}} \otimes_{j=1}^5 (|0\rangle + e^{-i2^{j-1}\phi}|1\rangle)$, which exhibits a ϕ -dependent phase modulation

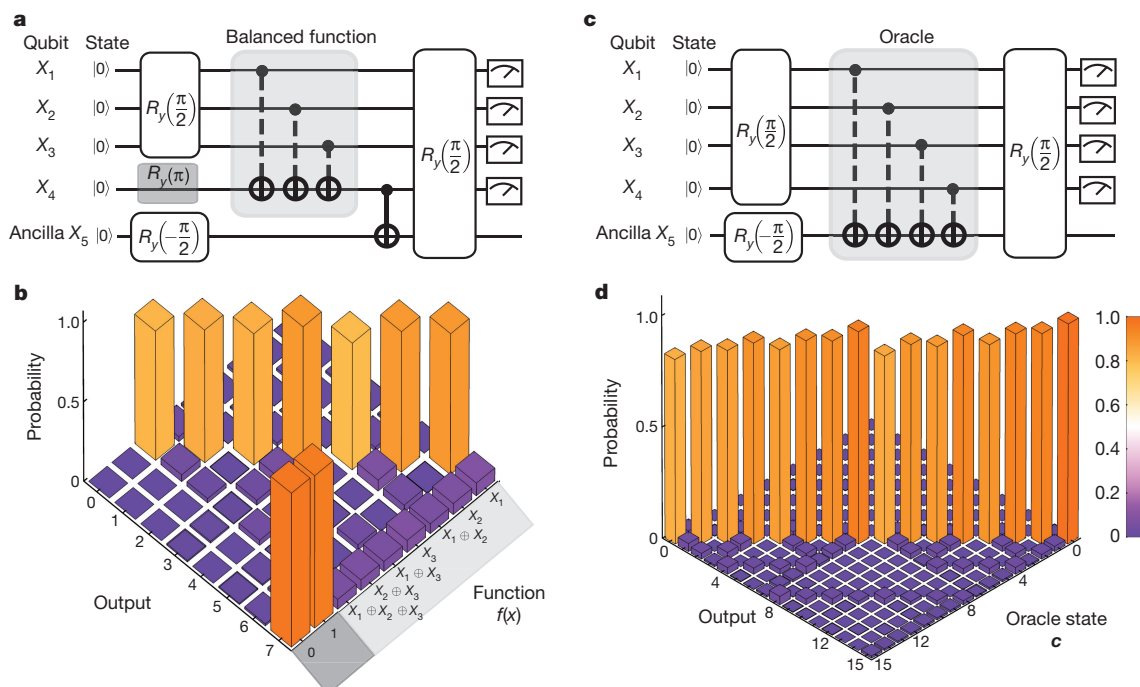


Figure 3 | Quantum algorithms. **a**, The Deutsch–Jozsa algorithm circuit on 5 ions. The oracle is implemented through gates shown in the shaded regions of the circuit. For balanced function oracles we apply each of the seven possible CNOT combinations, indicated in light grey. For the constant functions, we prepare $X_4 = 0$ or 1 as indicated in dark grey. **b**, Measured populations of the output state for various functions, conditioned upon measuring $X_4 = 1$. The two constant functions $f = 0$ and $f = 1$ are indicated in dark grey, and the seven balanced functions given by particular CNOT gate combinations are indicated in light grey. Measurement of the output $\{X_1 X_2 X_3\} = \{111\} = 7$ indicates a constant

function, while any other value (0–6) indicates a balanced function. **c**, The Bernstein–Vazirani algorithm circuit. The shaded region contains programmed CNOT gate combinations used to implement different oracle states **c**. **d**, Measured output population for various oracle states. The output is the inverted oracle state \bar{c} . Data represented in **b**, **d** are obtained by sampling over 20,000 experimental repetitions for each function or the oracle state **c** and the errors for the success probabilities in each case are statistical uncertainties within 1 s.d. The displayed probabilities are colour coded (key at right).

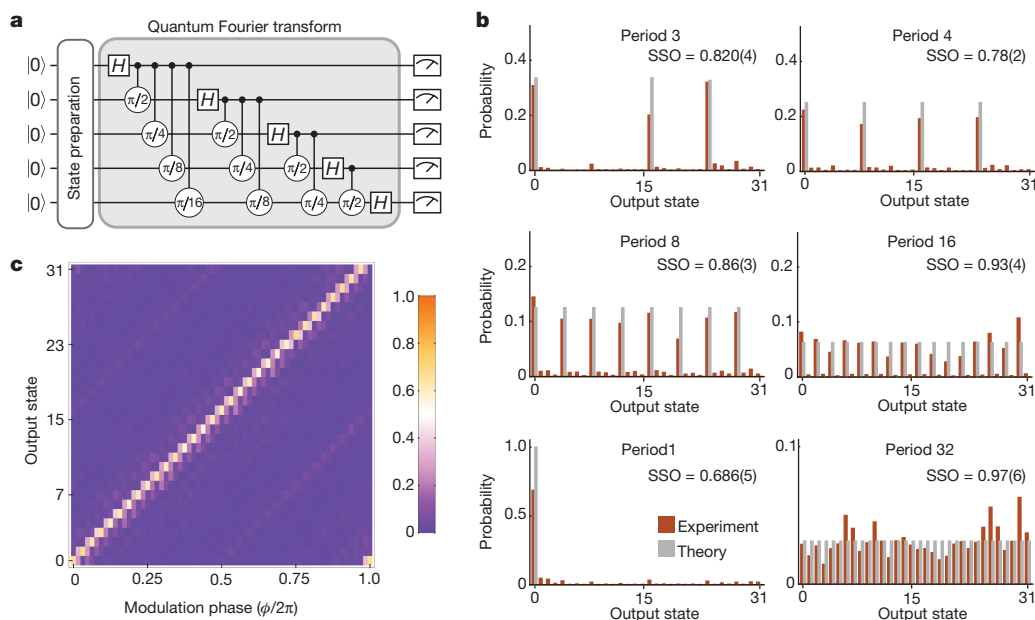


Figure 4 | Quantum Fourier transform protocol. **a**, Experimental sequence for implementation and verification of the quantum Fourier transform (QFT). ‘State preparation’ consists of single qubit rotations that create a phase and amplitude modulation of the coefficients $\{C_k\}$ of the input state $\sum_{k=0}^{31} C_k |k\rangle$. The shaded grey region contains a sequence of modular gates for implementing the QFT, which is then followed by a measurement of the register. **b**, Quantum period finding. Input states are prepared using single-qubit rotations to modulate the 32 state amplitudes with periods 1, 3, 4, 8, 16, and 32 (see Methods). The squared statistical

overlap (SSO)⁵ signifies the fidelity of the protocol where the error is a statistical estimate over 8,000 experimental repetitions. The grey and red bars represent populations calculated from theory and measured in the experiment, respectively. **c**, Quantum phase estimation using five measurement qubits. The plot shows populations in the output state that estimates the given phase modulation ϕ of the input state amplitudes $\{C_k\}$. Probabilities in the output state population are colour coded. We observe the correct value of the phase in each case with a probability > 0.6 . The experiment is repeated 8,000 times for each value of ϕ .

$C_k = \frac{1}{\sqrt{32}} e^{-ik\phi}$. We apply the QFT on this state to estimate ϕ by mapping its value onto populations of the output state, as shown in Fig. 4c. This is repeated for several cases where ϕ is incremented in steps of $2\pi/64$ over the range 0 to 2π . Values of ϕ that are integer multiples of $2\pi/32$ result in the output state $|32\phi/2\pi\rangle$. This is achieved with an average fidelity of 0.619(5). For non-integer values, the population is distributed between the nearest 5-bit approximate states¹⁴.

In our experiments, each algorithm fidelity is limited mainly by the native gate errors (<2%), which propagate into the standard logic gate errors (<5%) (see Methods). These errors are dominated by Raman beam imperfections and therefore can be reduced by mitigating Raman beam intensity noise²³ and individual addressing crosstalk (see Methods). Systematic shifts in the axes of the gate rotations accumulate due to unequal Stark shifts across the qubits, which result in algorithmic errors that depend upon the circuit structure. This type of error can be easily eliminated by feeding forward known shifts to the radio frequency of individual qubit control beams.

The algorithms presented here illustrate the computational flexibility provided by the trapped-ion quantum architecture. Within a single module, this system can be scaled to dozens of qubits by linearly increasing the number of radio-frequency controls and AOM- and PMT-channels at the hardware level. In software, the number of XX- and R-gate calibrations required to compile any logic gate scale as $\mathcal{O}(n^2)$. As more ions are added to the chain, the axial confinement must be weakened to maintain a linear crystal. This will slow down the XX-gate duration roughly as $n^{1.7}$, but the crosstalk is not expected to get worse (see Methods). Finally, implementing this architecture on multi-zone ion traps such as surface traps will provide further control over the connectivity of qubits through shuttling¹⁵ for scalable computation. This will also enable selective measurement of qubits that can be fed-forward classically to perform conditional operations in the module⁵ as required for fault-tolerant computing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 March; accepted 1 June 2016.

- Linden, N., Barjat, H. & Freeman, R. An implementation of the Deutsch-Jozsa algorithm on a three-qubit NMR quantum computer. *Chem. Phys. Lett.* **296**, 61–67 (1998).
- Vandersypen, L. M. K. *et al.* Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
- Gulde, S. *et al.* Implementation of the Deutsch-Jozsa algorithm on an ion-trap quantum computer. *Nature* **421**, 48–50 (2003).
- Brainis, E. *et al.* Fiber-optics implementation of the Deutsch-Jozsa and Bernstein-Vazirani quantum algorithms with three qubits. *Phys. Rev. Lett.* **90**, 157902 (2003).
- Chiaverini, J. *et al.* Implementation of the semiclassical quantum Fourier transform in a scalable system. *Science* **308**, 997–1000 (2005).
- Brickman, K.-A. *et al.* Implementation of Grover's quantum search algorithm in a scalable system. *Phys. Rev. A* **72**, 050306(R) (2005).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Shi, F. *et al.* Room-temperature implementation of the Deutsch-Jozsa algorithm with a single electronic spin in diamond. *Phys. Rev. Lett.* **105**, 040504 (2010).
- Martín-López, E. *et al.* Experimental realization of Shor's quantum factoring algorithm using qubit recycling. *Nat. Photon.* **6**, 773–776 (2012).
- Monz, T. *et al.* Realization of a scalable Shor algorithm. *Science* **351**, 1068–1070 (2016).
- Deutsch, D. & Jozsa, R. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).
- Bernstein, E. & Vazirani, U. Quantum complexity theory. *SIAM J. Comput.* **26**, 1411–1473 (1997).
- Shor, P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509 (1997).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 1st edn (Cambridge Univ. Press, 2002).
- Kielinski, D., Monroe, C. & Wineland, D. J. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
- Monroe, C. *et al.* Large scale modular quantum computer architecture with atomic memory and photonic interconnects. *Phys. Rev. A* **89**, 022317 (2014).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Mølmer, K. & Sørensen, A. Multipartite entanglement of hot trapped ions. *Phys. Rev. Lett.* **82**, 1835–1838 (1999).
- Barends, R. *et al.* Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature* **508**, 500–503 (2014).
- Hill, C. D. *et al.* A surface code quantum computer in silicon. *Sci. Adv.* **1**, e1500707 (2015).
- Gottesman, D. Fault-tolerant quantum computation with local gates. *J. Mod. Opt.* **47**, 333–345 (2000).
- Green, T. J. & Biercuk, M. J. Phase-modulated decoupling and error suppression in qubit-oscillator systems. *Phys. Rev. Lett.* **114**, 120502 (2015).
- Ballance, C. J. *et al.* Laser-driven quantum logic gates with precision beyond the fault-tolerant threshold. Preprint at <http://arxiv.org/abs/1512.04600> (2016).
- True Merrill, J. *et al.* Demonstration of integrated microscale optics in surface-electrode ion traps. *New J. Phys.* **13**, 103005 (2011).
- Gaebler, J. P. *et al.* High-fidelity universal gate set for $^9\text{Be}^+$ ion qubits. Preprint at <http://arxiv.org/abs/1604.00032> (2016).
- Choi, T. *et al.* Optimal quantum control of multimode couplings between trapped ion qubits for scalable entanglement. *Phys. Rev. Lett.* **112**, 190502 (2014).
- Olmschenk, S. *et al.* Manipulation and detection of a trapped Yb^+ hyperfine qubit. *Phys. Rev. A* **76**, 052314 (2007).
- Fisk, P. T. H., Sellars, M. J., Lawn, M. A. & Coles, C. Accurate measurement of the 12.6GHz “clock” transition in trapped $^{171}\text{Yb}^+$ ions. *IEEE Trans. Ultrasonics Ferroelectrics Frequency* **44**, 344–354 (1997).
- Hayes, D. *et al.* Entanglement of atomic qubits using an optical frequency comb. *Phys. Rev. Lett.* **104**, 140501 (2010).
- Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. Entanglement-free Heisenberg-limited phase estimation. *Nature* **450**, 393–396 (2007).

Acknowledgements We thank K. R. Brown, J. Kim, T. Choi, Z.-X. Gong, T. A. Manning, D. Maslov and C. Volin for discussions. This work was supported by the US Army Research Office with funds from the IARPA MQCO and LogiQ Programs, the Air Force Office of Scientific Research MURI on Quantum Measurement and Verification, and the National Science Foundation Physics Frontier Center at JQI.

Author Contributions S.D., N.M.L., C.F., K.A.L., K.W. and C.M. all contributed to the experimental design, construction, data collection and analysis of this experiment. All authors contributed to this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.D. (sdebath@umd.edu).

Reviewer Information *Nature* thanks S. Bartlett and T. Northup for their contribution to the peer review of this work.

METHODS

Experimental techniques. We use a linear radio-frequency Paul trap made of four segmented blade electrodes driven at 23.83 MHz where the transverse secular frequency of the trap is actively stabilized³¹. For measurement, state-dependent fluorescence is collected by a 0.38 numerical aperture objective that images ions with 0.55 μm resolution. For a single qubit, single-shot detection fidelities for states $|0\rangle$ and $|1\rangle$ are 99.74(3)% and 99.09(5)%, respectively. For $n = 5$ qubits, detection is degraded by signal crosstalk between PMT channels, and the average single-shot fidelity is 95.3(2)% for the 2^n states. For the population distributions measured in Figs 3 and 4 and the reported algorithm fidelities, multi-qubit detection is performed by signal-averaging the populations of all 2^n states over a few thousand experimental repetitions. In this way, detection and crosstalk errors are removed by decomposing the measurements into the known detector array response of all 32 possible qubit states. The individual addressing Raman beams are modulated using a multi-channel AOM (Model H-601 Series 32-Channel UV Acousto-Optic Modulator, PN: 66948-226460-G01, Harris Corporation) and focused down to a beam waist of approximately 1.5 μm at the ions. Addressing crosstalk between neighbouring ions due to Raman beam spillover is $<4\%$, which can be improved using higher resolution optics³².

As more ions are added to a chain, the ratio of axial-to-transverse confinement must be weakened to maintain a linear crystal ($\nu_z/\nu_x < 0.6n^{-0.86}$)³³. For constant transverse confinement, this means that the minimum ion spacing remains the same. However, this will slow the gates down. In our setup (for $n = 5$) two-qubit XX-gates for any ion pair $\{i, j\}$ have a duration of $\tau_g = 235 \mu\text{s}$, which depends on the spectral splitting of the transverse modes ($\tau_g \sim \nu_x/\nu_z^2 \sim n^{1.7}$). The XX-gate pulse shape is a 9-segment piecewise-constant Rabi frequency modulation $\{\Omega_{kij}\}$ (where $1 \leq k \leq 9$), which is implemented by modulating the global Raman beam. Optimized pulse shapes are calculated for each ion pair such that $\{\Omega_{kij}\}$ is within practical limits and the gate fidelity is maximized. The number of classical calculations to find the pulse shapes scales as $\mathcal{O}(n^2)$. The XX-gates are calibrated by setting the product of the laser intensities on the two qubits such that $\chi_{ij} = \pi/4$ (refs 26, 34–36). For CP gates that require other values of χ_{ij} , we scale the laser intensity accordingly. Single-qubit rotations are calibrated by measuring the Rabi frequency Ω_i of individual qubits. Single-qubit native R-gates have a duration of approximately $0.1\tau_g$.

Implementation of the Deutsch–Jozsa algorithm. The Deutsch–Jozsa algorithm is implemented by starting with an equal superposition of all classical input states to the function $f(x): \{0, 1, \dots, 7\} \rightarrow \{0, 1\}$. We prepare this by initializing all qubits to $|0\rangle$, followed by $R_y(\pi/2)$ rotations on the qubits in the control register $x = X_1X_2X_3$. Then we rotate the ancilla qubit X_5 using $R_y(-\pi/2)$. The resulting 5-qubit state is

$$|\psi\rangle_0 = \frac{1}{\sqrt{8}} \sum_{x=0}^7 |x\rangle_{123} \otimes |0\rangle_4 \otimes \frac{|0\rangle_5 - |1\rangle_5}{\sqrt{2}}$$

where x is the decimal representation of qubits $X_1X_2X_3$. Then we apply the function on the input superposition state such that the value is written to X_4 . The resulting state is

$$|\psi\rangle_1 = \frac{1}{\sqrt{8}} \sum_{x=0}^7 |x\rangle_{123} |f(x)\rangle_4 \otimes \frac{|0\rangle_5 - |1\rangle_5}{\sqrt{2}}$$

This is followed by a CNOT between the function register X_4 and the ancilla X_5 which provides a phase ‘kick-back’ to produce the state

$$|\psi\rangle_2 = \frac{1}{\sqrt{8}} \sum_{x=0}^7 (-1)^{f(x)} |x\rangle_{123} |f(x)\rangle_4 \otimes \frac{|0\rangle_5 - |1\rangle_5}{\sqrt{2}}$$

This is followed by a single-qubit rotation $R_y(\pi/2)$ on all qubits. Then we measure the first four qubits to reach the solution and ignore the ancilla qubit since it is not entangled with the other qubits. The state of qubits $X_1X_2X_3X_4$ before measurement can be written as

$$\begin{aligned} |\psi\rangle_3 &= \frac{1}{8} \sum_{y=0}^7 \sum_{x=0}^7 (-1)^{f(x)} (-1)^{\bar{y} \cdot x} |\bar{y}\rangle_{123} \otimes \frac{|1\rangle_4 + (-1)^{f(x)} |0\rangle_4}{\sqrt{2}} \\ &= C_{0000}|0000\rangle + C_{0001}|0001\rangle + \dots \\ &\dots + C_{1110}|1110\rangle + C_{1111}|1111\rangle \end{aligned} \quad (1)$$

where \bar{y} is the bit-wise inversion of y . If $f(x) = a$ is a constant function (with $a = \{0, 1\}$), the coefficients of the basis states $|1110\rangle$ and $|1111\rangle$ are

$$\begin{aligned} C_{1110} &= \frac{1}{8\sqrt{2}} (-1)^a \sum_{x=0}^7 (-1)^{000 \cdot x} = \frac{(-1)^a}{\sqrt{2}} \\ C_{1111} &= \frac{1}{8\sqrt{2}} \sum_{x=0}^7 (-1)^{000 \cdot x} = \frac{1}{\sqrt{2}} \end{aligned}$$

If $f(x)$ is a balanced function, then the coefficients are

$$\begin{aligned} C_{1110} &= \frac{1}{8\sqrt{2}} \sum_{x=0}^7 (-1)^{000 \cdot x} (-1)^{f(x)} (-1)^{f(x)} = \frac{1}{\sqrt{2}} \\ C_{1111} &= \frac{1}{8\sqrt{2}} \sum_{x=0}^7 (-1)^{000 \cdot x} (-1)^{f(x)} = 0 \end{aligned}$$

Here we use the property that $f(x) = 0$ for exactly half of the values of x and 1 for the rest. Conditioned upon $X_4 = 1$, there is unit probability of measuring $X_1, X_2, X_3 = 111$ for a constant function and 0 probability of measuring the same outcome when the function is balanced. In equation (1), note that the probability of measuring $X_4 = 1$ is 0.5 irrespective of the number of qubits in the input (control) register of the function.

Native single- and two-qubit rotations. Native single-qubit operations $R_\phi(\theta)$ are rotations of the Bloch vector by an angle θ about an axis on the equator of the Bloch sphere, where ϕ is the angle between this rotation axis and the X axis. The single-qubit operator is

$$R_\phi(\theta) = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & -i\sin\left(\frac{\theta}{2}\right)e^{-i\phi} \\ -i\sin\left(\frac{\theta}{2}\right)e^{i\phi} & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}$$

The standard X and Y rotations used in the composite gates are simply $R_x(\theta) = R_0(\theta)$ and $R_y(\theta) = R_{\pi/2}(\theta)$.

Native two-qubit XX-gates are performed by invoking a $\sigma_x \sigma_x$ -Ising interaction between qubits i and j , which is mediated through the coupling of the qubits to the collective transverse motional modes of the ion chain. The resulting two-qubit entangling rotation $XX(\chi_{ij})$ depends on the geometric phase χ_{ij} , which is the integrated Ising interaction and can be varied by changing the Raman beam intensity. The sign of the geometric phase $\alpha_{ij} = \text{sgn}(\chi_{ij})$ depends on how ions i and j couple to the common transverse motional modes. The XX-gate operator is

$$XX(\chi_{ij}) = \begin{bmatrix} \cos(\chi_{ij}) & 0 & 0 & -i\sin(\chi_{ij}) \\ 0 & \cos(\chi_{ij}) & -i\sin(\chi_{ij}) & 0 \\ 0 & -i\sin(\chi_{ij}) & \cos(\chi_{ij}) & 0 \\ -i\sin(\chi_{ij}) & 0 & 0 & \cos(\chi_{ij}) \end{bmatrix}$$

In this experiment, $\alpha_{12}, \alpha_{45}, \alpha_{14}, \alpha_{25}, \alpha_{35}, \alpha_{23}, \alpha_{34} = +1$ and $\alpha_{15}, \alpha_{25}, \alpha_{13} = -1$.

Composite gate fidelity. Controlled-NOT (CNOT) gates are performed between all ion pairs and characterized in the following way. We perform the CNOT gate on all four classical input states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ and measure the fidelity from the population of the desired output state. The average fidelity of a CNOT on each ion pair is shown in Extended Data Table 1.

Controlled-phase (CP) gates are performed between all ion pairs and characterized by using a sequence of gates. We first initialize the qubits in the state $\frac{1}{\sqrt{2}}(|1\rangle(|0\rangle + |1\rangle))$, where the first qubit is the control qubit and the second qubit is the target qubit. This is followed by a conditional phase gate $CP(\theta)$ that creates the state $\frac{1}{\sqrt{2}}(|1\rangle(|0\rangle + e^{i\theta}|1\rangle))$. A final rotation $R_x(\pi/2)$ on the target qubit projects the conditional phase θ onto the population of the target qubit as $P(|1\rangle) = (1 - \sin\theta)/2$. This is shown in Extended Data Figure 1.

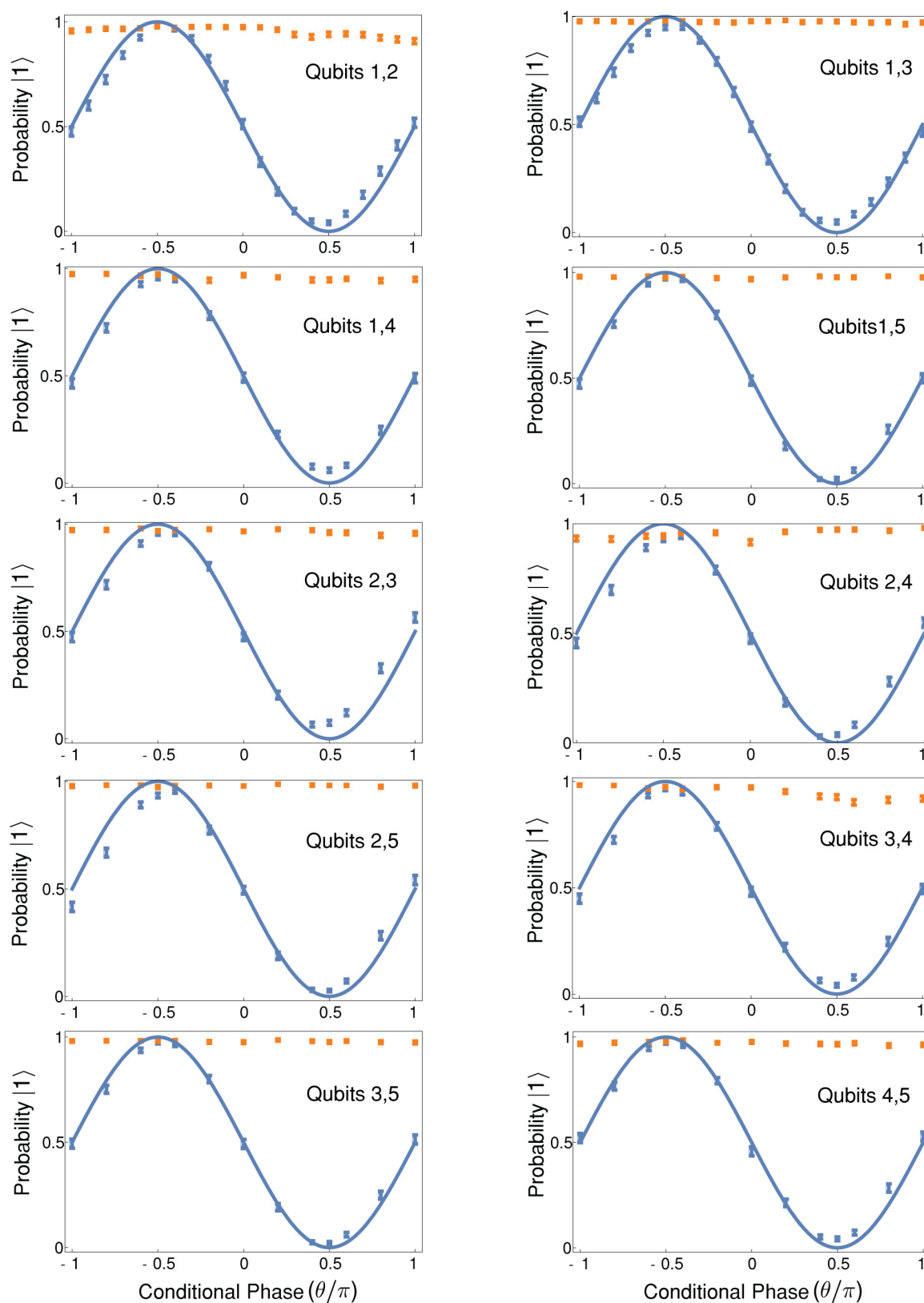
We measure the fidelity of the CP gates at conditional phases $\theta = \pm\pi/2$, which correspond to the maximum and minimum values of θ , respectively, which are used in a coherent QFT or QFT⁻¹. At these values of θ , where the geometric phase $\chi_{ij} = \pi/4$, the XX-gates are most sensitive to laser intensity fluctuations, which leads to maximum errors. This is evident from the data shown in Extended Data Figure 1, where a maximum deviation of the target qubit from the ideal output state occurs at $\pm\pi/2$. Therefore, the fidelity measure at these values is a lower bound on the CP gate fidelity. The fidelity is obtained by measuring the populations in

the $|10\rangle$ and $|11\rangle$ states for $\theta = +\pi/2$ and $\theta = -\pi/2$, respectively. Extended Data Table 2 shows the fidelities of all CP gates.

QFT state preparation. For the period-finding experiment, an amplitude or phase modulation is created in the coefficients C_k of the input state $\sum_{k=0}^{31} C_k |k\rangle$ using individual single-qubit rotations. Extended Data Table 3 shows the input states for various measured periodicities.

Sample size. No statistical methods were used to predetermine sample size.

31. Johnson, K. G. *et al.* Active stabilization of ion trap radiofrequency potentials. *Rev. Sci. Instrum.* **87**, 053110 (2016).
32. Crain, S., Mount, E., Baek, S. & Kim, J. Individual addressing of trapped $^{171}\text{Yb}^+$ ion qubits using a microelectromechanical systems-based beam steering system. *Appl. Phys. Lett.* **105**, 181115 (2014).
33. Schiffer, J. P. Phase transitions in anisotropically confined ionic crystals. *Phys. Rev. Lett.* **70**, 818–821 (1993).
34. Zhu, S.-L., Monroe, C. & Duan, L.-M. Trapped ion quantum computation with transverse phonon modes. *Phys. Rev. Lett.* **97**, 050505 (2006).
35. Solano, E., de Matos Filho, R. L. & Zagury, N. Deterministic Bell states and measurement of the motional state of two trapped ions. *Phys. Rev. A* **59**, R2539–R2543 (1999).
36. Milburn, G. J., Schneider, S. & James, D. F. V. Ion trap quantum computing with warm ions. *Fortschr. Phys.* **48**, 801–810 (2000).



Extended Data Figure 1 | Controlled-phase gate. Shown is the performance of the controlled-phase (CP) gate between control (red) and target (blue) qubit for different qubit-pairs. The control qubit is prepared in the state $|1\rangle$ which remains unchanged during the gate. Solid blue lines

indicate the theoretical probability of measuring the target qubit in $|1\rangle$ whereas the data points show experimental data. Error bars are statistical, indicating a 95% confidence interval for 2,000 experimental repetitions.

Extended Data Table 1 | Controlled-NOT gate fidelities

Ion pair	Fidelity (%)	Ion pair	Fidelity (%)
1,2	96.4(6)	2,4	98.5(7)
1,3	97.6(7)	2,5	96.8(7)
1,4	95.9(7)	3,4	96.6(5)
1,5	97.9(5)	3,5	97.6(6)
2,3	95.6(6)	4,5	97.2(5)

Gate fidelity is obtained by performing CNOT gates on all possible pairs of ions ('ion pair') in a chain of five qubits.

Extended Data Table 2 | Controlled-phase gate fidelities

Ion pair	$\theta = \frac{\pi}{2}$, fidelity (%)	$\theta = -\frac{\pi}{2}$, fidelity (%)
1,2	91.1(6)	96.1(4)
1,3	93.6(5)	93.3(6)
1,4	91.6(6)	93.3(6)
1,5	95.9(4)	95.3(3)
2,3	90.7(6)	93.2(5)
2,4	94.2(5)	90.8(6)
2,5	95.8(4)	91.7(6)
3,4	91.0(6)	94.7(5)
3,5	96.0(4)	96.0(4)
4,5	93.5(6)	95.8(4)

Gate fidelity is obtained by performing CP gates on all possible pairs of ions ('ion pair') in a chain of five qubits for conditional phases $\theta = \pm\pi/2$.

Extended Data Table 3 | Input states in QFT-period finding

Input state	Period
$\frac{1}{\sqrt{32}}(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)$	1
$\frac{1}{\sqrt{32}}(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+e^{i6.2\pi/16} 1\rangle)(0\rangle+i 1\rangle)$	3
$\frac{1}{\sqrt{8}}(0\rangle+ 1\rangle)(0\rangle+ 1\rangle)(0\rangle+ 1\rangle) 11\rangle$	4
$\frac{1}{2}(0\rangle+ 1\rangle)(0\rangle+ 1\rangle) 111\rangle$	8
$\frac{1}{\sqrt{2}}(0\rangle+ 1\rangle) 1111\rangle$	16
$ 11111\rangle$	32

See Methods for details.

Carbon-based tribofilms from lubricating oils

Ali Erdemir¹, Giovanni Ramirez¹, Osman L. Eryilmaz¹, Badri Narayanan², Yifeng Liao^{1†}, Ganesh Kamath² & Subramanian K. R. S. Sankaranarayanan²

Moving mechanical interfaces are commonly lubricated and separated by a combination of fluid films and solid ‘tribofilms’, which together ensure easy slippage and long wear life¹. The efficacy of the fluid film is governed by the viscosity of the base oil in the lubricant; the efficacy of the solid tribofilm, which is produced as a result of sliding contact between moving parts, relies upon the effectiveness of the lubricant’s anti-wear additive (typically zinc dialkyldithiophosphate)². Minimizing friction and wear continues to be a challenge, and recent efforts have focused on enhancing the anti-friction and anti-wear properties of lubricants by incorporating inorganic nanoparticles and ionic liquids^{3,4}. Here, we describe the *in operando* formation of carbon-based tribofilms via dissociative extraction from base-oil molecules on catalytically active, sliding nanometre-scale crystalline surfaces, enabling base oils to provide not only the fluid but also the solid tribofilm. We study nanocrystalline catalytic coatings composed of nitrides of either molybdenum or vanadium, containing either copper or nickel catalysts, respectively. Structurally, the resulting tribofilms are similar to diamond-like carbon⁵. Ball-on-disk tests at contact pressures of 1.3 gigapascals reveal that these tribofilms nearly eliminate wear, and provide lower friction than tribofilms formed with zinc dialkyldithiophosphate. Reactive and *ab initio* molecular-dynamics simulations show that the catalytic action of the coatings facilitates dehydrogenation of linear olefins in the lubricating oil and random scission of their carbon–carbon backbones; the products recombine to nucleate and grow a compact, amorphous lubricating tribofilm.

Transportation vehicles account for about 19% of the world’s energy consumption and some 23% of total greenhouse-gas emissions every year⁶. With ever-increasing mobility, these numbers will undoubtedly surge and present challenges for sustainable transportation^{7,8}. So far, new efficiency and emission standards imposed on vehicles have been the main driving force behind the development of cleaner and more fuel-efficient lubricants over the years⁹. Most efforts have involved lowering the viscosity of base oils, and exploring new ways to replace zinc dialkyldithiophosphate (ZDDP) and other additives that contain sulfated ash, phosphorous and sulphur (SAPS) with more environmentally friendly alternatives, including inorganic nanoparticles, ionic liquids and coatings^{3,4,10,11}. It is desirable to further reduce the use of environmentally harmful additives that give rise to adverse emissions, without compromising on performance in terms of protection against friction and wear.

Here, we report the design and synthesis of a new breed of catalytically active nanocomposite coatings. These coatings are made of the nitrides of transition metals such as molybdenum and vanadium (90–95% by atomic weight), together with metal catalysts such as copper (5–10% by atomic weight)—producing, for example, the MoN_x–Cu coating described below and the VN–Cu coating described in the Supplementary Information. Figure 1a shows a cross-sectional transmission electron microscopy (TEM) image of the MoN_x–Cu coating, which was deposited on martensitic chrome steel substrates

(specifically, AISI 52100 steel), which have a thickness of about 600 nm. The high-resolution transmission electron microscopy (HRTEM) image in Fig. 1b reveals a very dense and nanocrystalline structure, composed of copper and MoN_x grains of 5–10 nm in diameter. Figure 1c shows composite Cu–K and Mo–K edge energy-dispersive X-ray spectroscopy (EDS) images in cross-section, which confirm the presence of copper-rich clusters throughout the film (Extended Data Figs 1 and 2 show detailed HRTEM and X-ray diffraction (XRD) analyses of the nanocomposite coatings). X-ray photoelectron spectroscopy (XPS) revealed peaks belonging to the key constituents of the coating (molybdenum, copper and nitrogen) and to some surface contaminants (mostly oxides; Extended Data Fig. 3). We estimated the amount of copper in the film to be about 5% by XPS. Nanomechanical characterization of the MoN_x–Cu coating revealed hardness values of about 20 GPa and an elastic modulus of around 235 GPa (Extended Data Fig. 4).

We determined the friction and wear behaviour of the nanocomposite with a ball-on-disk test pair, in which a stationary ball is pressed against a rotating disk and both are coated in a lubricant

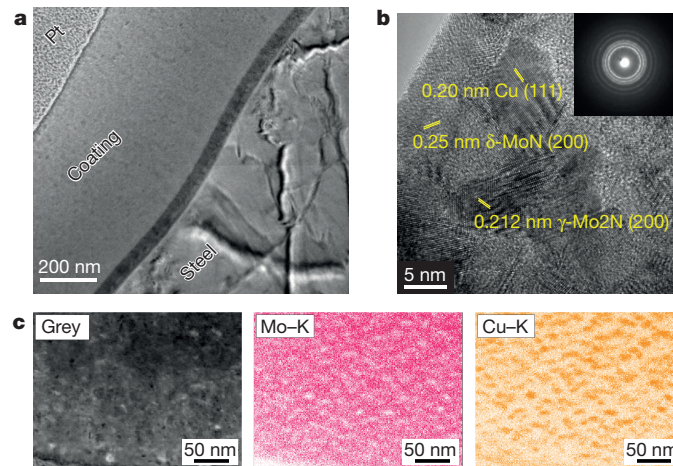


Figure 1 | Structure of the MoN_x–Cu nanocomposite coating. **a**, General cross-sectional TEM image of a MoN_x–Cu sample prepared with a focused ion beam, confirming a dense, compact coating of approximately 600 nm thickness. The thin, dark-looking layer above the steel is the molybdenum bonding layer between the coating and the substrate (60 nm thick). The layer above the coating is the protective platinum film, needed for focused ion-beam milling. **b**, High-resolution TEM image taken from a region near the top surface, showing the highly nanocrystalline structure of the MoN_x–Cu film with crystal size of less than 10 nm. The inset shows the diffraction pattern of the polycrystalline structure. δ-MoN_x is hexagonal and γ-MoN_x is cubic (see Extended Data Figs 1 and 2). **c**, Cu–K and Mo–K edge EDS elemental mapping of the film, confirming the existence of copper clusters (lighter regions in Mo–K and darker regions in Cu–K images) uniformly distributed within the film structure (EDS sampling depth is estimated to be between 1 and 2 μm). The ‘grey’ image is a STEM image.

¹Energy Systems Division, 9700 South Cass Avenue, Argonne National Laboratory, Argonne, Illinois 60439, USA. ²Center for Nanoscale Materials, 9700 South Cass Avenue, Argonne National Laboratory, Argonne, Illinois 60439, USA. [†]Present address: Dow Corning Corporation, 2200 West Salzburg Road, Midland, Minnesota 48642, USA.

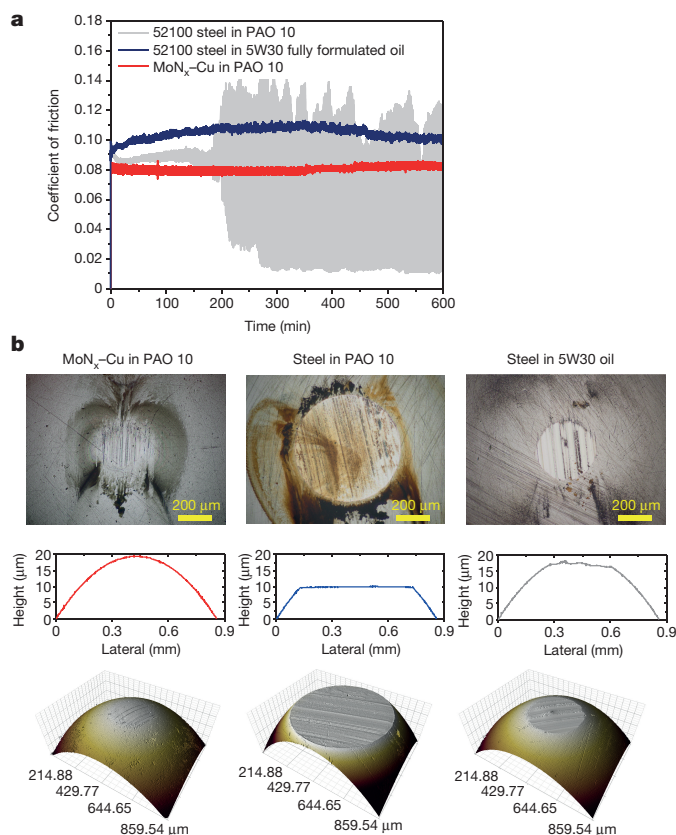


Figure 2 | Friction and wear behaviour of the MoN_x-Cu-coated steel ball and disk in PAO 10 oil, and comparison with uncoated steel in PAO 10 or fully formulated 5W30 oils. a, Comparison of friction coefficients. **b**, Top, optical micrographs obtained from rubbing ball surfaces, showing the condition and difference in wear scar size for the three tests (see also Fig. 2a). Below, line scans and three-dimensional profiles reveal the extent of wear damage more clearly. There is unmeasurable wear for the ball coated with MoN_x-Cu, while the steel ball tested in PAO 10 oil shows the highest wear loss, revealing a very flat and polished wear scar after the 10-hour test (here, contact pressure was reduced to 70 MPa owing to the formation of a flat wear scar). The steel ball tested in formulated 5W30 oil shows a much smaller wear scar than the steel ball tested in PAO 10 (contact pressure was reduced to 282 MPa for the test with 5W30 oil).

(see Supplementary Methods). Figure 2a compares the friction coefficient of a MoN_x-Cu-coated test pair with those of uncoated 52100-steel test pairs in pure poly- α -olefin (PAO) 10 and fully formulated 5W30 oil (which contains ZDDP and a suite of other additives). Despite the severe contact pressure (1.3 GPa) between ball and disk, the friction coefficient of MoN_x-Cu is about 0.08 and very steady throughout the test. The amount of wear is limited to a few scratches; otherwise, the highly spherical nature of the original ball surface is well preserved (Fig. 2b). Closer inspection reveals the formation of a tribofilm during the sliding process. This film accumulates around the trailing edge of the contact spot as a black debris layer. Right on the top of the contact spot, where sliding has occurred, the tribofilm appears much thinner and is difficult to discern by optical microscopy, but we were able to confirm its presence by time-of-flight secondary-ion mass spectrometry (TOF-SIMS) (Extended Data Fig. 5).

The friction coefficient of the uncoated 52100 test pair in PAO oil is about 0.09 initially, but after about 200 minutes it becomes very unsteady, fluctuating between 0.02 and 0.12. This is most likely due to the generation and accumulation of many wear-related debris particles at the sliding-contact interface, giving rise to high frictional instability and severe wear losses (Fig. 2b). In fact, subsequent calculations of wear-induced loss for the steel ball indicate a total volumetric wear of $1.34 \times 10^{-12} \text{ m}^3$ ($1.05 \times 10^{-5} \text{ g}$) (Extended Data Fig. 6). The friction

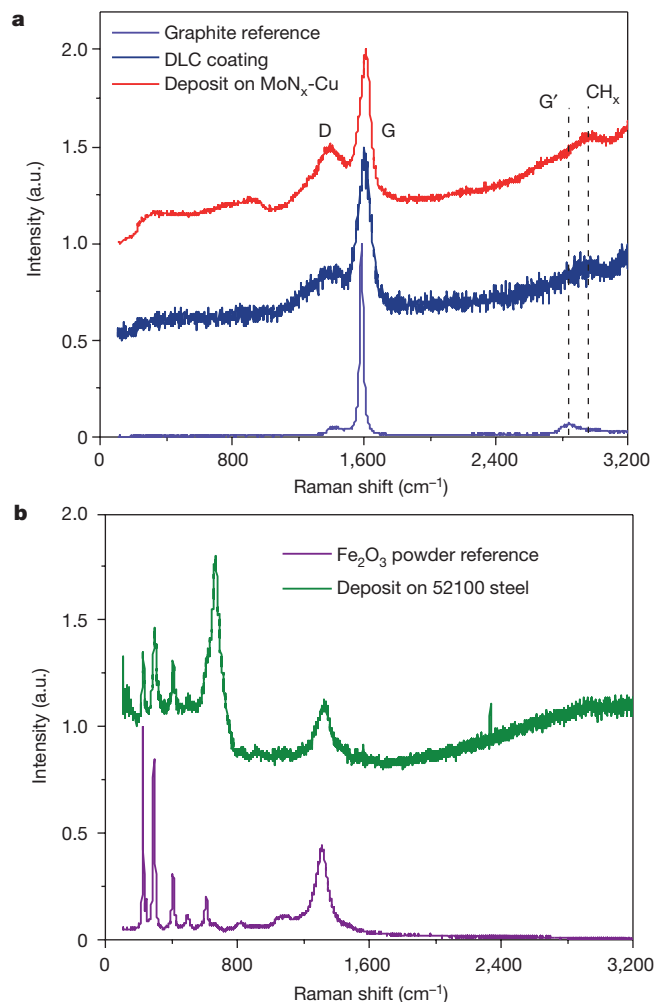


Figure 3 | Raman spectra of the tribofilms produced during tests of coated and uncoated steel balls in pure PAO 10. a, UV Raman (325 nm) spectra of: the black debris patch or the deposits found in and around the edge of the wear scar of a ball coated with MoN_x-Cu; the solid, diamond-like carbon (DLC) film produced by magnetron sputtering; and pyrolytic graphite (used as a reference). The D and G bands of the carbon deposits are clearly observed in positions very close to those of the graphite reference; the predominant intensity of the G peak and the width of the peaks show in a qualitative way that the carbon deposit (or tribofilm) is structurally amorphous (as is DLC) with a high fraction of sp^2 bonding, suggesting that the carbon deposit produced during rubbing is graphitic. **b**, Visible Raman spectra of the brown patch found around the wear scar of uncoated 52100 steel. These spectra are analogous to those obtained from Fe₂O₃ powder (used as a reference), with multiple Raman bands that correspond to one another.

coefficient of the steel test pair in a fully formulated 5W30 oil is around 0.1 (Fig. 2a); the ball wear ($1.26 \times 10^{-13} \text{ m}^3$; $9.86 \times 10^{-7} \text{ g}$) is much lower than that of the steel ball tested in PAO but higher than that of the MoN_x-Cu coated ball in PAO (Extended Data Fig. 6). The extent of wear damage on the flat samples is commensurate with the damage found on the ball surfaces (Extended Data Fig. 7).

Elucidating the chemical nature of the debris particles found in and around the wear scars is important for understanding the mechanisms that are responsible for the tribological events taking place during sliding. Accordingly, we have analysed the chemistry of the debris particles using confocal Raman spectroscopy and TEM. Figure 3a is a composite of three Raman spectra (obtained with a 325-nm laser): one from the black debris layer seen in Fig. 2b, one from a diamond-like carbon (DLC) coating (produced by magnetron sputtering in argon/methane plasma), and one from pyrolytic graphite (as a reference). The spectrum from the black debris layer on the MoN_x-Cu coated

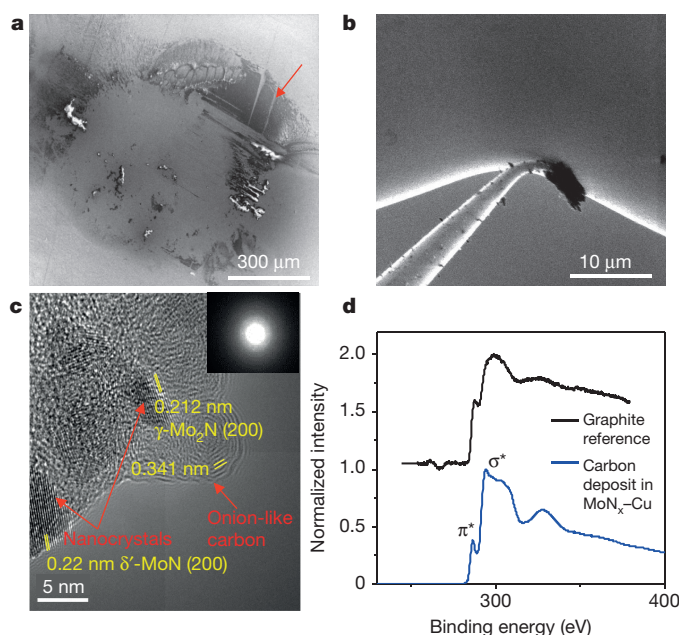


Figure 4 | TEM sample preparation and results. **a**, SEM image of the area around the wear scar on a steel ball coated with MoN_x-Cu; carbon-rich debris was collected from the area indicated by the red arrow. **b**, Collection of the carbon-rich debris (the tribofilm), using a tungsten tip in the focused ion beam with the help of a micromanipulator. **c**, High-resolution TEM near the edge of the scooped debris. The presence of a large amorphous area, with some nanocrystals and onion-like carbon, is revealed; the electron diffraction (inset) confirms the amorphous nature of the tribofilm. **d**, EELS spectrum (deconvoluted with the zero loss) of the tribofilm, in comparison with the spectrum for a highly oriented pyrolytic graphite (reference), showing that the film is made of mainly graphitic carbon.

ball is similar to that from the DLC film, as it shows one-to-one correspondence of the D (at around 1,395 cm⁻¹) and G (1,600–1,610 cm⁻¹) bands. The spectrum from pyrolytic graphite indicates that the black debris around the wear scar of the MoN_x-Cu coated ball is made mostly of *sp*²-bonded carbon (as implicit from the broadness and high intensity of the G band). We attribute the broad feature at around 2,900 cm⁻¹ to CH_x stretching in both the DLC and the debris layer¹². The shift of the G band above 1,600 cm⁻¹ (in comparison with graphite, at 1,583 cm⁻¹) suggests a high degree of disorder in the debris layer and DLC, as well as the presence of some *sp*² chains¹².

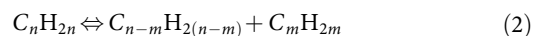
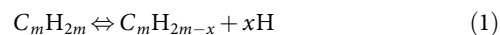
As for the brownish debris around the wear scar of the 52100 ball tested in PAO oil, the Raman analyses (Fig. 3b) confirm that this was made mostly of iron oxides (several peaks from the debris match those from the Fe₂O₃ reference). Obviously, during rubbing, the dominant wear mechanism was oxidative wear¹³, and the oxide-based debris was thus controlling the wear and friction of this test pair.

In Fig. 4, we present our findings from TEM of the black debris layers shown in Fig. 2b. Briefly, we used a tungsten tip to transfer part of the debris from the top right section (denoted with a red arrow) to a copper post of an OmniProbe lift-out grid (Fig. 4a, b). The edge of the debris had thin regions that were electron transparent. Figure 4c shows an HRTEM image, which reveals that the debris is amorphous, but that some nanocrystals are also present. These crystals are about 5 nm in size and from the MoN_x (on the basis of the lattice fringe/*d* spacing; Extended Data Fig. 8). The EDS of the debris layer reinforces this assertion, revealing strong peaks for molybdenum (Extended Data Fig. 9). (Because of the background from the TEM column and the copper grid, the amount of copper is not quantifiable from the EDS.) The presence of such crystalline MoN_x particles within the debris is not surprising, as the MoN_x-Cu coating experienced a certain degree of abrasion (see the scratches in Fig. 2b). During sliding, these crystalline MoN_x particles became an integral part of the debris layer, accumulating

around the edges of the wear scars in Fig. 2b. The electron-diffraction pattern in Fig. 4c and the electron energy-loss spectra (EELS) in Fig. 4d show that most of the debris seen in Figs 2b and 4a is composed of amorphous carbon that has high *sp*² character, as manifested by a strong π^* peak at 285 electronvolts (eV). On the basis of the π^* and σ^* peaks, we calculated the fraction of *sp*²-bonded carbon atoms in the debris to be about 82%, consistent with the Raman spectra (Fig. 3a). The Raman spectra also showed dominant G-peak positions corresponding to those of graphite; the broadness of these peaks (compared with those from graphite) is due to the smaller cluster size, the cluster distribution, and the influence of stress in the debris¹². Onion-like carbon¹⁴ was also observed by HRTEM (Fig. 4c), suggesting the transformation of oil molecules to other carbon allotropes at the sliding-contact interface of MoN_x-Cu.

Our results provide clear evidence that the MoN_x-Cu coating can transform PAO molecules (hydrogenated alkenes) into carbon-based tribofilms at the sliding interface. These films are similar to films of hydrogenated amorphous carbon with high *sp*² bonding. Our tribofilms were able to nearly eliminate wear and to provide lower friction than those resulting from 5W30 oil, despite the severe contact pressure (Fig. 2b). Mechanistically it is well known that, under lubricated sliding conditions, a protective tribofilm is derived from the ZDDP in formulated oils^{2,15} mainly because mechanical rubbing under high contact pressure and shear stress generates heat^{13,16}, plastic deformation¹⁷, and structural defects¹⁸—all of which increase the chemical activity of the rubbing interface. We expect that these events occur more favourably with PAO- or mineral-oil-lubricated MoN_x-Cu or VN-Cu coatings than with uncoated surfaces, because the activation energy for dissociative extraction of carbon-based tribofilms from the long-chain hydrocarbon molecules of lubricating oils is lowered.

To understand the atomistic mechanism of tribofilm formation, we carried out extensive *ab initio* molecular dynamics (AIMD) and reactive molecular-dynamics (RMD) simulations, starting from an initial configuration composed of linear olefin chains sandwiched between the sliding tribological interfaces (see Supplementary Methods). Both simulations were performed at 1,000 K, which is representative of the asperity-level flash temperatures of typical tribology experiments^{13,16}. Our AIMD simulations illustrate the elementary steps involved in the catalytic dissociation of olefin (Fig. 5a–c). A similar pathway is predicted by our RMD simulations, which also capture the nucleation and growth of the dissociated olefin into a carbon tribofilm (Fig. 5d–f). In both simulations we observe that, on non-carbide-forming surfaces such as copper, the olefins catalytically convert to tribofilms that are reminiscent of hydrogenated amorphous carbon, following the atomistic pathway shown in Fig. 5a–c. Initially, the olefin molecules are uniformly distributed (Fig. 5d). Then, under the sliding action at the Cu-olefin interface, the olefins degrade via two competing steps: first, dissociation of C–H bonds near the Cu surface, leading to dehydrogenated chains (equation (1) and Fig. 5b, e); and second, random scission of backbone C–C bonds to form shorter hydrocarbon fragments (equation (2) and Fig. 5c, f).



Note that it is easier to break a C–C bond than a C–H bond owing to its lower bond-dissociation energy (C–C: ~347 kJ mol⁻¹; C–H: ~414 kJ mol⁻¹; refs 19, 20). However, in the presence of metal catalysts such as nickel and copper, dehydrogenation is facilitated^{21–23} via a three-centred transition state, in which a C–H bond bridges across the metal until the bond breaks, after which metal–H and metal–C bonds form (see Supplementary Methods for details of charge-transfer dynamics, chemical-bond evolution and the electronic origin of catalytic activity on different metal surfaces). The released free hydrogen either gets adsorbed onto the metal surface and diffuses into its bulk, or recombines at the sliding interface to form H₂ molecules. Meanwhile, the dissociated

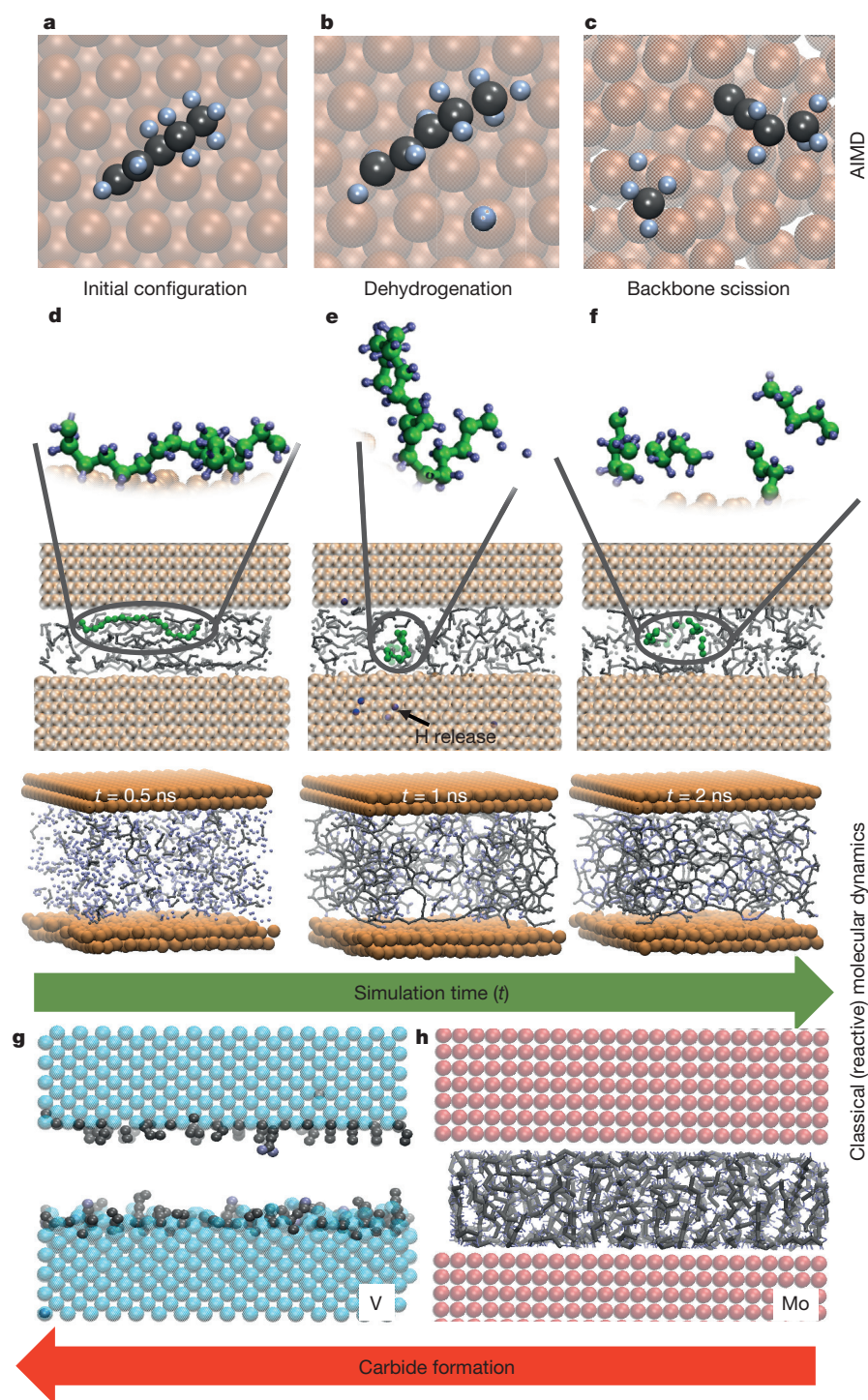


Figure 5 | Atomistic mechanism of tribofilm formation by $\text{MoN}_x\text{-Cu}$, deduced from *ab initio* and reactive molecular-dynamics simulations. **a–c**, AIMD simulations illustrate the catalytic action of copper, dehydrogenating and breaking linear olefins into shorter-chain hydrocarbons (see Supplementary Video 5). **d–f**, A similar reaction pathway is predicted by our reactive molecular-dynamics simulations (see Supplementary Video 1). The end result is that dehydrogenated short-chain hydrocarbons recombine to nucleate, and grow into, a compact amorphous carbon tribofilm. Simulations suggest that this tribofilm mechanism is suppressed on surfaces where carbide formation is thermodynamically favoured (see Supplementary Video 2 and Supplementary Information for the evolution of various chemical bonds). **g, h**, Snapshots from reactive molecular-dynamics simulations show that two carbide-forming surfaces (vanadium and molybdenum) do not, to all intents and purposes, form tribofilms. Although a sliding molybdenum surface does (like copper) show olefin degradation, the catalytic activity at 1,000 K is much lower and the resulting kinetics of tribofilm formation is very sluggish (see Supplementary Videos 3, 4, 6 and 7). Detailed first-principles calculations show that MoN and VN have much reduced catalytic activity compared with copper and nickel (see Supplementary Video 8 and Supplementary Information for the electronic origin of differences in the catalytic behaviour of various metals and nitrides).

short-chain dehydrogenated hydrocarbons recombine to form graphitic tribofilm (Fig. 5d–f; see Supplementary Information for details of the temporal evolution of C–H, C–C, metal–C, metal–H and H–H bonds). An analysis of the final configuration shows that about 80% of the carbon atoms are sp^2 -hybridized, consistent with our experiments.

It has been shown that graphitic tribofilms can even form *in vivo*, on the rubbing surfaces of metal-on-metal hip replacements (which are made of cobalt, chrome and molybdenum)²⁴. Proteins are suspected to be the main carbon source in these tribofilms, which result from the catalytic nature of the cobalt and molybdenum in the metal-on-metal alloy. In our case, the extraction of tribofilm from PAO molecules results from the catalytic nature of the composite $\text{MoN}_x\text{-Cu}$ coating (while metal nitrides without a catalyst exhibit much reduced catalytic activity, as suggested by our simulations; see Supplementary Information).

In summary, we have shown that catalytically active coatings enable base oils to provide solid tribofilms that protect sliding surfaces against friction and wear. Our concept could be extended to the redesign of bulk materials to incorporate catalytic metals, leading to the formation of similar carbon-based tribofilms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 May 2015; accepted 7 June 2016.

1. Bowden, F. P., Gregory, J. N. & Tabor, D. Lubrication of metal surfaces by fatty acids. *Nature* **156**, 97–101 (1945).
2. Spikes, H. The history and mechanisms of ZDDP. *Tribol. Lett.* **17**, 469–489 (2004).

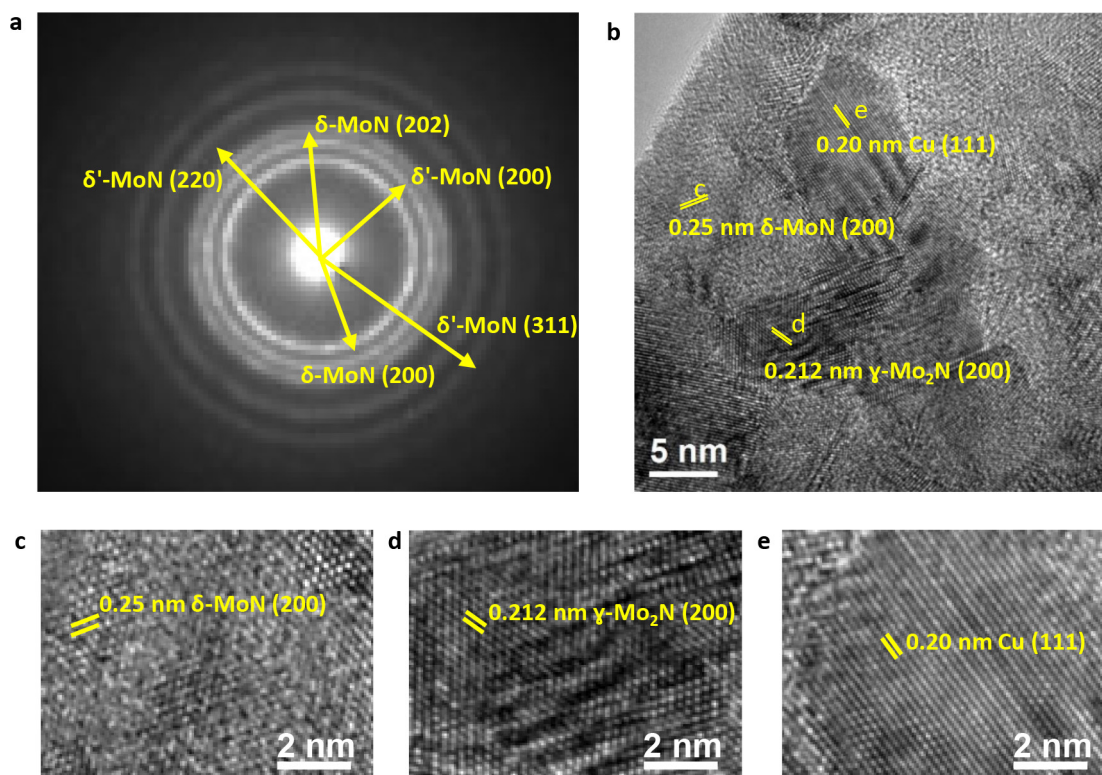
3. Tenne, R. Recent advances in the research of inorganic nanotubes and fullerene-like nanoparticles. *Front. Phys.* **9**, 370–377 (2014).
4. Qu, J. *et al.* Synergistic effects between phosphonium-alkylphosphate ionic liquids and zinc dialkyldithiophosphate (ZDDP) as lubricant additives. *Adv. Mater.* **27**, 4767–4774 (2015).
5. Erdemir, A. & Donnet, C. Tribology of diamond-like carbon films: recent progress and future prospects. *J. Phys. D* **39**, R311–R327 (2006).
6. World Energy Council *Global Transport Scenarios 2050* (World Energy Council, 2012).
7. Chu, S. & Majumdar, A. Opportunities and challenges for a sustainable energy future. *Nature* **488**, 294–303 (2012).
8. Holmberg, K., Andersson, P. & Erdemir, A. Global energy consumption due to friction in passenger cars. *Tribol. Int.* **47**, 221–234 (2012).
9. Tung, S. C. & McMillan, M. L. Automotive tribology overview of current advances and challenges for the future. *Tribol. Int.* **37**, 517–536 (2004).
10. Spikes, H. Low- and zero-sulphated ash, phosphorus and sulphur anti-wear additives for engine oils. *Lubr. Sci.* **20**, 103–136 (2008).
11. Cha, S. C. & Erdemir, A. *Coating Technology for Vehicle Applications* (Springer International Publishing, 2015).
12. Ferrari, A. & Robertson, J. Resonant Raman spectroscopy of disordered, amorphous, and diamondlike carbon. *Phys. Rev. B* **64**, 075414 (2001).
13. Lim, S. C. & Ashby, M. F. Wear-mechanism maps. *Acta Metall.* **35**, 1–24 (1987).
14. Kroto, H. W. Carbon onions introduce new flavour to fullerene studies. *Nature* **359**, 670–671 (1992).
15. Gosvami, N. N. *et al.* Mechanisms of antiwear tribofilm growth revealed in situ by single-asperity sliding contacts. *Science* **348**, 102–106 (2015).
16. Bowden, F. P. & Ridler, K. E. W. Physical properties of surfaces. III. The surface temperature of sliding metals the temperature of lubricated surfaces. *Proc. R. Soc. A* **154**, 640–656 (1936).
17. Alpas, T., Hu, H. & Zhang, J. Plastic deformation and damage accumulation below the worn surfaces. *Wear* **162–164**, 188–195 (1993).
18. Johnson, K. L. Contact mechanics and the wear of metals. *Wear* **190**, 162–170 (1995).
19. David, D. J. & Misra, A. *Relating Materials Properties to Structure with MATPROP Software: Handbook and Software for Polymer Calculations and Materials Properties* (CRC Press, 2001).
20. Huheey, J. E., Keiter, E. A. & Keiter, R. L. *Inorganic Chemistry: Principles of Structure and Reactivity* (HarperCollins, 1993).
21. Sattler, J. J. H. B., Ruiz-Martinez, J., Santillan-Jimenez, E. & Weckhuysen, B. M. Catalytic dehydrogenation of light alkanes on metals and metal oxides. *Chem. Rev.* **114**, 10613–10653 (2014).
22. Kang, D. B. & Anderson, A. B. Theoretical interpretation of the cyclohexane → benzene reaction on the Pt(III) surface. *J. Am. Chem. Soc.* **107**, 7858–7861 (1985).
23. Ceamanos, J., Mastral, J., Millera, A. & Aldea, M. Kinetics of pyrolysis of high density polyethylene. Comparison of isothermal and dynamic experiments. *J. Anal. Appl. Pyrolysis* **65**, 93–110 (2002).
24. Liao, Y. *et al.* Graphitic tribological layers in metal-on-metal hip replacements. *Science* **334**, 1687–1690 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Work at Argonne National Laboratory was supported by the US Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Vehicle Technologies and Advanced Manufacturing Offices under contract DE-AC02-06CH11357. Use of the Center for Nanoscale Materials was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357. This research also used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the US Department of Energy under contract DE-AC02-06CH11357. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract DE-AC02-05CH11231. XPS and TOF-SIMS measurements were carried out in the Frederick Seitz Materials Research Laboratory Central Research Facilities, University of Illinois. We thank R. Haasch for support with the XPS measurements and T. Spila for support with the TOF-SIMS measurements. We thank D. Gosztola for help with the Raman measurements at the Center of Nanoscale Materials, Argonne National Laboratory. We also thank S. Rodil, O. Depablos and R. Mirabal for help with the GIXRD measurements at Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México. We thank E. A. Bardasz of Zual Associate in Lubrication for helpful discussions, and W. Connacher and M. McKinnon for their help in setting up the experiments.

Author Contributions A.E. conceived the idea and directed the research project. O.L.E. is the co-principal investigator and, together with G.R., produced and optimized the nanocomposite coatings. G.R. carried out the XRD, XPS, nano-indentation and Raman measurements and analysis of the data resulting from those techniques. G.R. and O.L.E. carried out the tribological experiments. G.R., A.E. and O.L.E. analysed the data obtained during the tribological characterization. Y.L. carried out the HRTEM and FIB preparation and imaging as well as the analysis of the EELS data. G.R. performed the phase analyses of HRTEM images and diffraction patterns. S.K.R.S. guided the simulation effort. B.N. performed all the first-principles calculations and reactive molecular-dynamics simulations. B.N., G.K. and S.K.R.S. performed all the related data analysis. A.E. composed the manuscript with input from all authors.

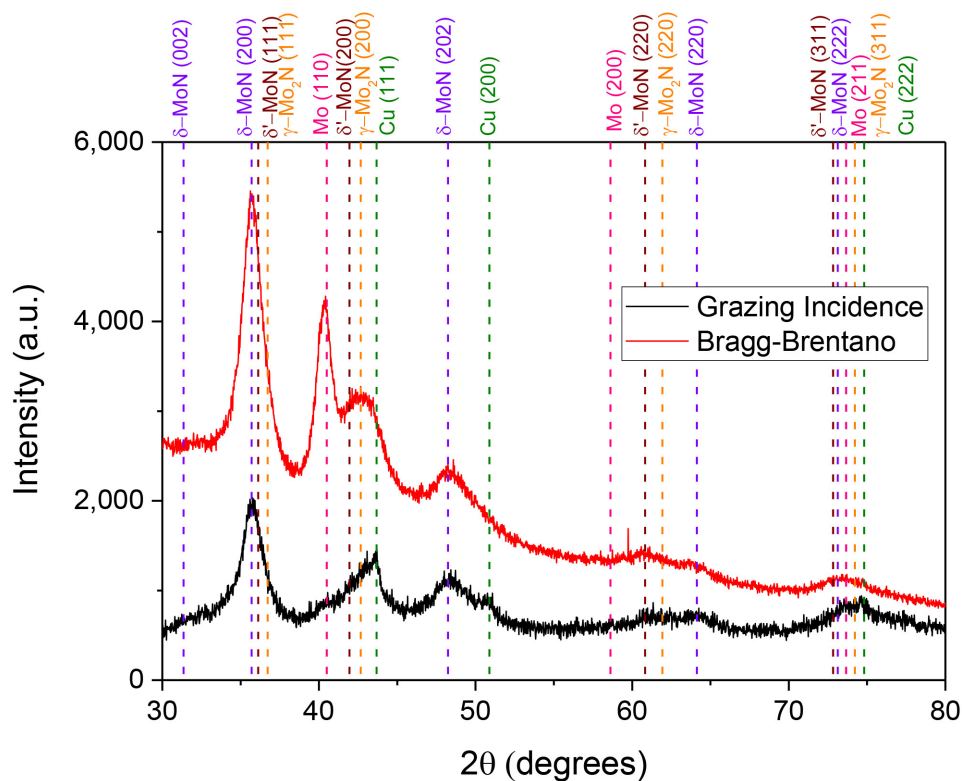
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.E. (erdemir@anl.gov).



Extended Data Figure 1 | HRTEM analysis of the MoN_x-Cu nanocomposite coating. Measurements of *d* spacing were performed using the DiffTools suite in DigitalMicrograph software²⁵. **a**, Diffraction pattern, showing the presence of two phases of MoN_x that have been reported in the Inorganic Crystal Structure Database (ICSD; <http://icsd.fiz-karlsruhe.de/>): hexagonal δ-MoN (ICSD accession number 99452) and

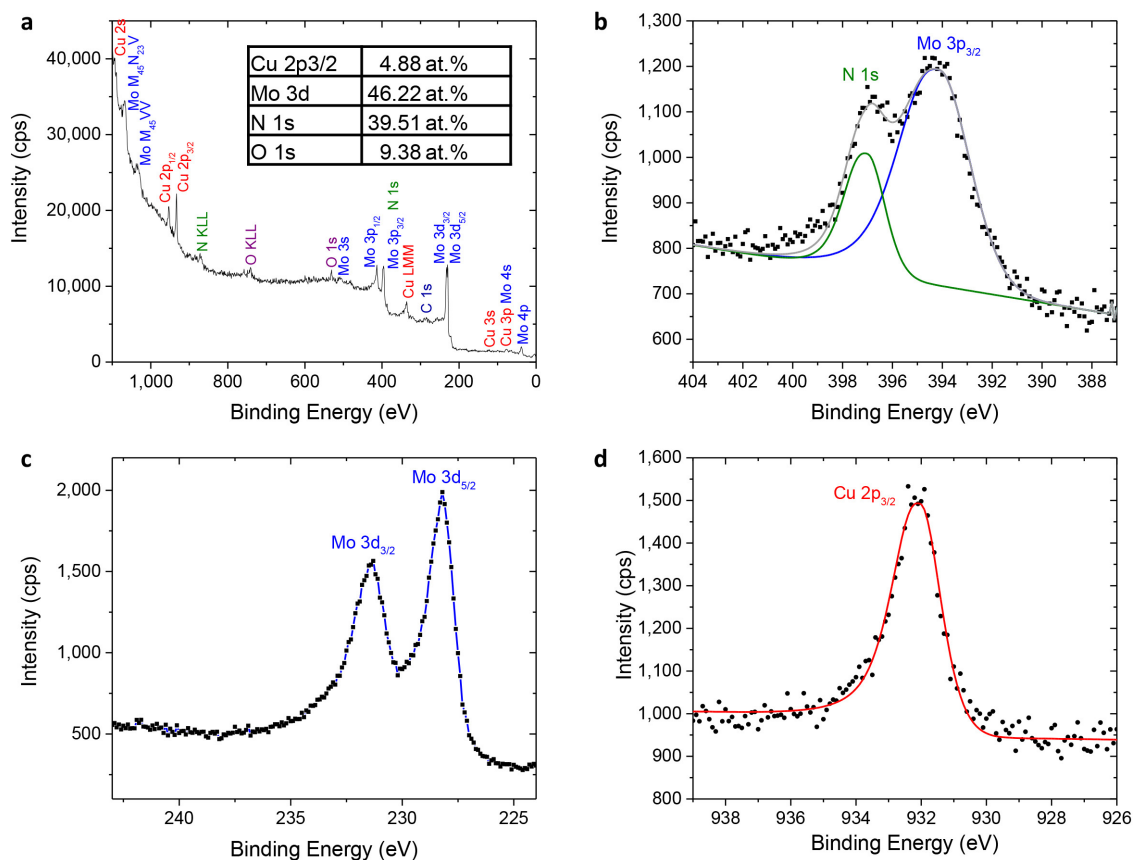
cubic δ'-MoN (ICSD accession number 159439). **b**, Measurements of the *d* spacing taken directly from the HRTEM image show the presence of nanocrystals of copper (ICSD 64699) in the film. **c**, δ-MoN found in the nanocomposite coating. **d**, γ-Mo₂N in the coating. **e**, Copper nanocrystal close to the surface. The numbers in parentheses in the figure are the Miller indices of crystallographic planes.

25. Mitchell, D. R. G. DiffTools: electron diffraction software tools for DigitalMicrograph. *Microsc. Res. Tech.* **71**, 588–593 (2008).
26. Kazmanli, M. K., Ürgen, M. & Cakir, F. Effect of nitrogen pressure, bias voltage and substrate temperature on the phase structure of Mo-N coatings produced by cathodic arc PVD. *Surf. Coat. Tech.* **167**, 77–82 (2003).
27. Oliver, W. C. & Pharr, G. M. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **7**, 1564–1583 (1992).
28. Hsu, W. K. *et al.* Electrolytic formation of carbon nanostructures. *Chem. Phys. Lett.* **262**, 161–166 (1996).
29. Qiao, Z., Li, J., Zhao, N., Shi, C. & Nash, P. Graphitization and microstructure transformation of nanodiamond to onion-like carbon. *Scr. Mater.* **54**, 225–229 (2006).



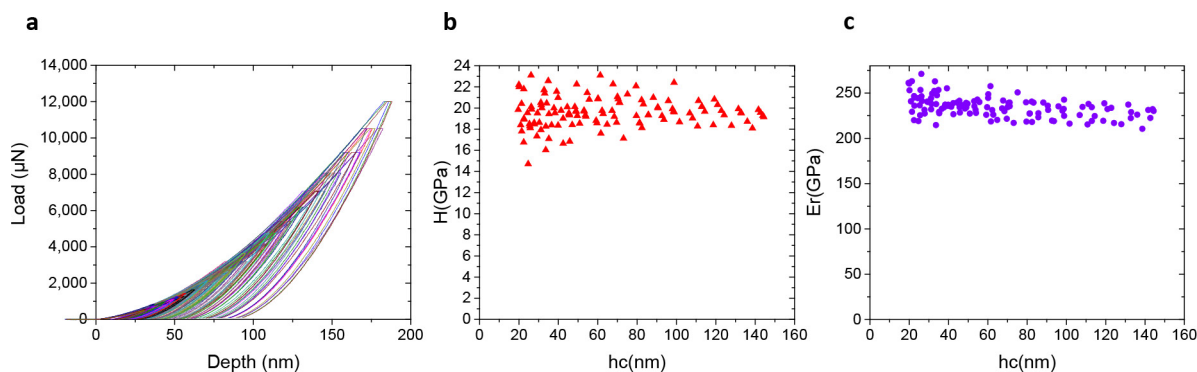
Extended Data Figure 2 | X-ray diffraction patterns of the $\text{MoN}_x\text{-Cu}$ coating. These spectra show the presence of copper, the molybdenum bonding layer and three different phases of MoN_x . The possible MoN phases that can be produced by physical vapour deposition have been well studied²⁶, but identifying such phases by XRD is difficult because the diffraction peaks can overlap. Here, grazing incidence XRD could not recognize the molybdenum bonding layer (between substrate and coating) that is identified with the peak at 40.5° in the Bragg–Brentano measurement, and which corresponds to the (110) plane of molybdenum (ICSD accession number 52267). The presence of the hexagonal $\delta\text{-MoN}$, cubic $\gamma\text{-Mo}_2\text{N}$ and cubic $\delta'\text{-MoN}$ in the $\text{MoN}_x\text{-Cu}$ nanocomposite coating was confirmed using the relevant peaks. The lattice parameters were calculated on the basis of the d spacing of the peaks. The peaks located at 35.7° , 48.24° and 64.14° correspond to the (200), (202) and (220) planes,

respectively, of $\delta\text{-MoN}$ (ICSD accession number 99452); lattice parameters are $a = 5.8035 \text{ \AA}$ and $c = 5.7006 \text{ \AA}$. The peak at 42.67° is well associated with the (200) plane of $\gamma\text{-Mo}_2\text{N}$ (ICSD accession number 158843); this peak is generally the one used to identify the presence of this phase²⁶; the lattice parameter is $a = 4.2345 \text{ \AA}$. The cubic $\delta'\text{-MoN}$ (ICSD accession number 159439) was found via the low intensity peak at 60.11° ; the lattice parameter is $a = 4.35 \text{ \AA}$. Copper is more evident on the grazing incidence XRD pattern (owing to the greater surface sensitivity compared with the Bragg–Brentano method and the possibility of observing planes that are different to those that are perpendicular to the surface). The peaks located at 43.68° , 50.88° and 74.82° are associated exclusively with the (111), (200) and (220) planes, respectively, of copper (ICSD accession number 64699) ($a = 3.5864 \text{ \AA}$).



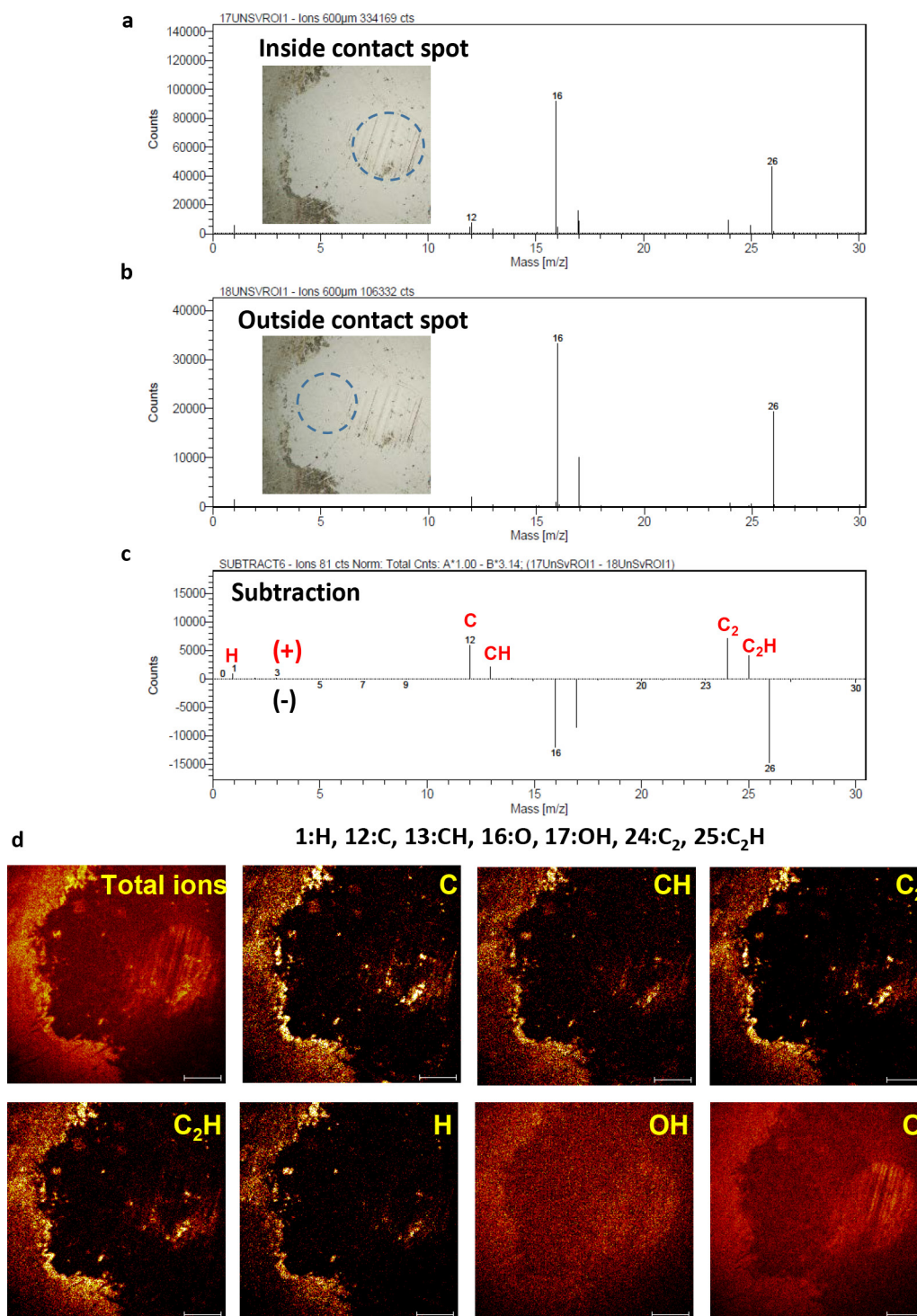
Extended Data Figure 3 | X-ray photoelectron spectroscopy of the MoN_x-Cu coating. **a**, Survey spectrum showing the presence of copper, molybdenum, nitrogen and oxygen in the film. The composition was calculated on the basis of the high-resolution peaks of the molybdenum 3d (doublet), oxygen 1s, nitrogen 1s and copper 2p_{3/2} orbitals. Intensity is

in counts per second. **b**, High-resolution spectra showing the peaks corresponding to the molybdenum 3p_{3/2} and nitrogen 1s orbitals. **c**, High-resolution spectrum of the molybdenum 3d orbital. **d**, High-resolution spectrum showing the copper 2p_{3/2} orbital.



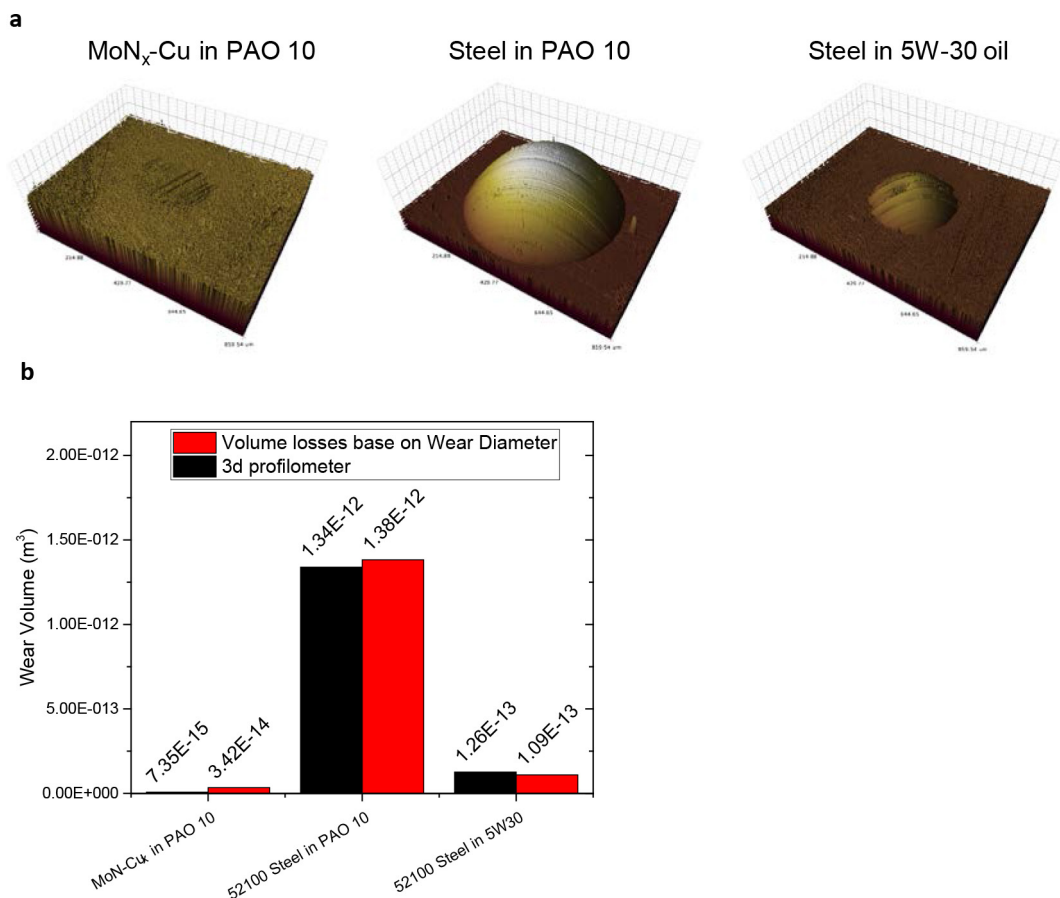
Extended Data Figure 4 | Nanoindentation analyses of the MoN_x-Cu coating. **a**, Loading/unloading curves for the measurements, using loads of between 500 μN and 12,000 μN. The 'depth' on the *x*-axis refers to the penetration depth of the diamond indentation tip. Different loads were used to evaluate the hardness (*H*) and elastic modulus (*E_r*) as a function of the contact depth (*h_c*), in order to avoid any influence of the substrate's

mechanical properties. The Oliver–Pharr method²⁷ was used to calculate the hardness and the elastic modulus of the coating. **b**, Hardness as a function of penetration. The hardness did not decrease with penetration, and was 19.6 ± 1.4 GPa. **c**, The elastic modulus of the coating was 234.7 ± 11.4 GPa across a range of indentations.



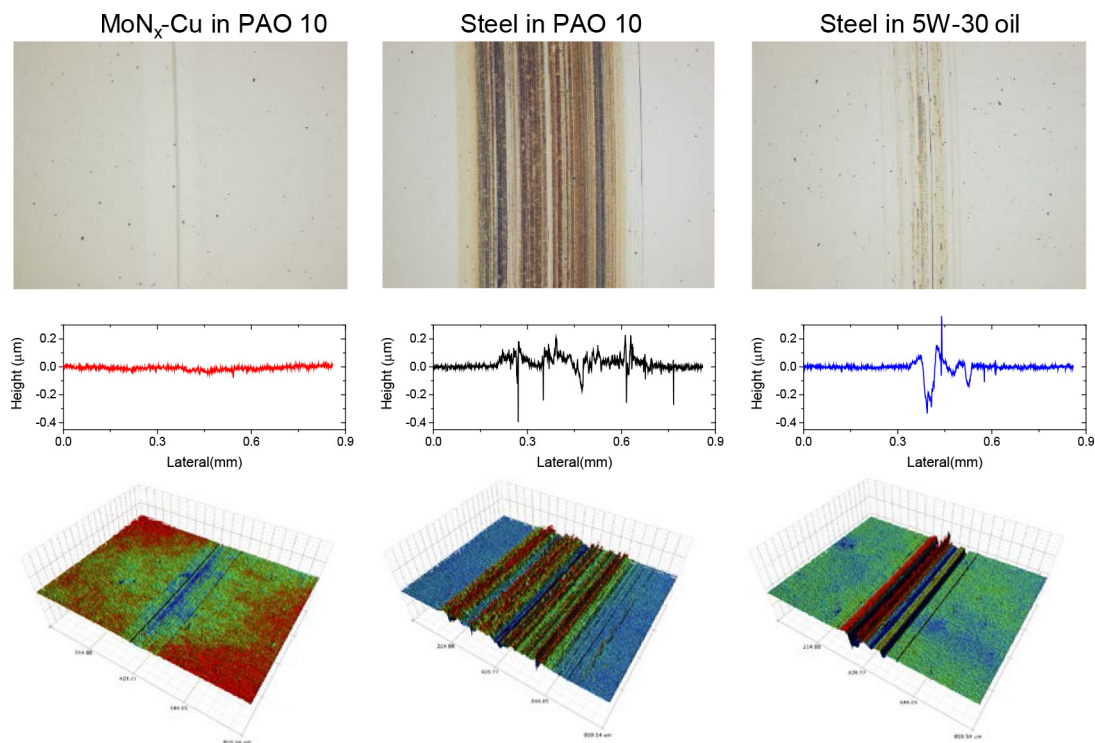
Extended Data Figure 5 | TOF-SIMS spectra and mapping of the contact region. See Supplementary Methods for experimental details. **a**, Spectrum of material from inside the contact spot (denoted with a circle). **b**, Spectrum of material from outside the contact spot (denoted with a circle). **c**, Spectrum after subtraction of the spectra in **a** and **b**, verifying the presence of carbon and other hydrocarbon fractions on

the contact spot (subtraction has been done after normalization of each spectrum according to total ion counts). **d**, Two-dimensional TOF-SIMS images from the contact spot and surrounding regions, confirming the presence of carbon and hydrocarbon fractions in the contact spot. The strong carbonaceous signals in the far outside section of the perimeter are due to the carbon-rich debris layers shown in Figs 2b and 4a.



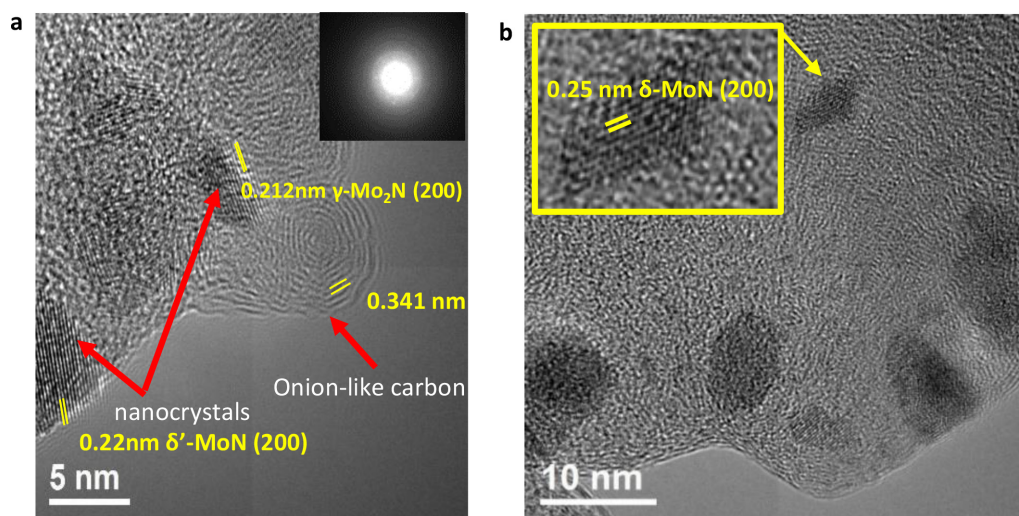
Extended Data Figure 6 | Wear volumes, calculated using three-dimensional optical profilometry. **a**, Three-dimensional representation of the volume loss of three steel balls (one coated with MoN_x-Cu and two uncoated, in different oils) on the basis of flattening of the spherical caps of the steel balls. There is very little wear, with only a few scratches, on the ball coated with MoN_x-Cu and tested in PAO 10.

b, Wear volume calculations, based on the profilometry in **a**, and compared with geometrical calculations based on the wear scar diameter. The wear volumes calculated using the two techniques are comparable, except in the case of MoN_x-Cu, where the diameter represents only the polished Hertzian contact area. 7.35E-15 is 7.35×10^{-15} , and so on.



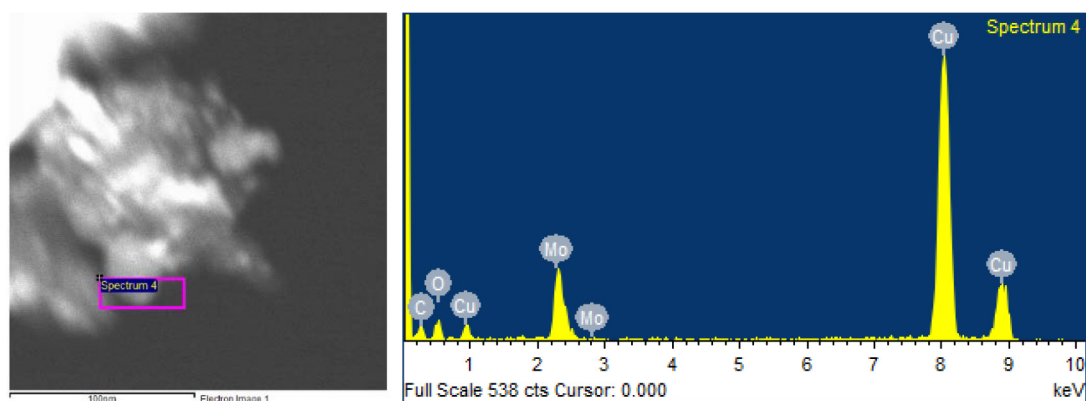
Extended Data Figure 7 | Wear produced on the flat surface of the disk rubbing against the balls. The optical micrographs at the top showing the physical condition and difference in wear track size in the three tests. The line scans and three-dimensional images below show the extent of

wear damage more clearly. There is unmeasurable wear for the MoN_x-Cu-coated flat surface. The steel surface tested in PAO 10 shows the highest wear loss, with a wider wear track and deep scratches; the flat surface tested in 5W30 oil shows a much smaller and narrower wear track.



Extended Data Figure 8 | Detailed analysis of HRTEM images of the debris layer of MoN_x-Cu. **a**, HRTEM image showing nanocrystals of δ' -MoN and γ -Mo₂N embedded in an amorphous carbon matrix. There were also onion-like carbons in many of the regions examined. The

d -orbital spacing of 0.341 nm is a typical value for this form of carbon structure^{28,29}. **b**, HRTEM image showing the amorphous matrix with one crystalline domain (yellow box), which corresponds to δ -MoN.



Extended Data Figure 9 | EDS of the debris layer of MoN_x-Cu tested in PAO 10. The spectrum shows the presence of molybdenum on the tribofilm. Copper cannot be quantified using this method owing the background from the TEM column. On the left is a TEM image showing the area of interest from which the spectrum was generated.

Evidence for climate change in the satellite cloud record

Joel R. Norris¹, Robert J. Allen², Amato T. Evan¹, Mark D. Zelinka³, Christopher W. O'Dell⁴ & Stephen A. Klein³

Clouds substantially affect Earth's energy budget by reflecting solar radiation back to space and by restricting emission of thermal radiation to space¹. They are perhaps the largest uncertainty in our understanding of climate change, owing to disagreement among climate models and observational datasets over what cloud changes have occurred during recent decades and will occur in response to global warming^{2,3}. This is because observational systems originally designed for monitoring weather have lacked sufficient stability to detect cloud changes reliably over decades unless they have been corrected to remove artefacts^{4,5}. Here we show that several independent, empirically corrected satellite records exhibit large-scale patterns of cloud change between the 1980s and the 2000s that are similar to those produced by model simulations of climate with recent historical external radiative forcing. Observed and simulated cloud change patterns are consistent with poleward retreat of mid-latitude storm tracks, expansion of subtropical dry zones, and increasing height of the highest cloud tops at all latitudes. The primary drivers of these cloud changes appear to be increasing greenhouse gas concentrations and a recovery from volcanic radiative cooling. These results indicate that the cloud changes most consistently predicted by global climate models are currently occurring in nature.

The International Satellite Cloud Climatology Project (ISCCP) dataset and the Extended Pathfinder Atmospheres (PATMOS-x) dataset are the two longest satellite records of cloudiness^{6,7}. The datasets consist of cloud retrievals made by multiple weather satellites over several decades, and in their original form, the long-term records suffer from spurious variability related to changes in satellite orbit, instrument calibration, and other factors^{4,5,8}. Previous studies using these datasets to investigate the cloud response to global warming have obtained inconclusive results owing to the dominating presence of artefacts^{9–11}. Here we use corrected versions of ISCCP and PATMOS-x from which spurious variability has been removed by empirically subtracting all cloud variability resembling an artefact⁵. Since one major artefact appears as coherent spurious cloud changes across the entire area viewed by a satellite, the correction procedure unfortunately also removes any real cloud variability at near-global scales, thus precluding examination of global mean cloud changes. Instead, we examined large-scale patterns of observed cloud change for consistency with patterns projected by global climate models to occur with climate change^{10,12}. For corroboration of the corrected ISCCP and PATMOS-x records, we additionally investigated the change in albedo observed in the 1980s by the Earth Radiation Budget Satellite (ERBS)¹³ and in the 2000s by the Clouds and Earth's Radiant Energy System (CERES) satellite instruments^{14,15}, and changes in the ocean-only liquid water path from the Multi-Sensor Advanced Climatology of Liquid Water Path (MAC-LWP) dataset¹⁶.

Figure 1a displays the spatial distribution of trends during the period 1983–2009 for the averaged ISCCP and PATMOS-x total cloud amount, and Fig. 1b displays the spatial distribution of differences between the

2002–2014 CERES albedo and the 1985–1989 ERBS albedo. All observational records agree that cloud amount and albedo increased over the northwest Indian Ocean, the northwest and southwest tropical Pacific Ocean, and north of the Equator in the Pacific and Atlantic oceans. Cloud amount and albedo decreased over mid-latitude oceans in both hemispheres (especially over the North Atlantic), over the southeast Indian Ocean, and in a northwest-to-southeast line stretching across the central tropical South Pacific. MAC-LWP exhibits a similar trend pattern in liquid water path during the period 1988–2014 (Extended Data Fig. 1a). Shifting the start or end time of trend calculation by several years has little impact on the spatial pattern of change. Similar patterns occur for differences in total cloud amount and albedo between the periods 1985–1989 and 2003–2009, during which ISCCP, PATMOS-x, and ERBS/CERES completely overlap (Extended Data Fig. 2).

Are the observed cloud changes solely a manifestation of natural internal variability or are they also a response to external radiative forcing of the climate system? We addressed this question by examining simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5) multi-model dataset¹⁷. Historical simulations included anthropogenic greenhouse gas concentrations, ozone, land-use changes, anthropogenic aerosols, volcanic aerosols, and solar output and thus represent our best estimate of the climate response to recent external radiative forcing (Extended Data Table 1). Figure 1c displays the spatial distribution of trends in ensemble mean modelled total cloud amount during the 27-year period 1983–2009 for all radiative forcings (ALL). Observations and models exhibit widespread agreement on which areas have increasing and decreasing cloud amount (Fig. 1d). Table 1 lists the spatial correlation between observed and simulated cloud trends.

Could natural internal variability alone produce the correlation between the observed and simulated cloud trend patterns? We assessed the likelihood of this outcome by examining cloud trends during 27-year periods from CMIP5 preindustrial simulations without external radiative forcing (Extended Data Table 2). Calculating the spatial correlation between the ensemble mean ALL trend pattern and the trend pattern of each 27-year preindustrial period generates a probability distribution of correlation values arising purely from natural internal variability (Extended Data Fig. 3). We found that no 27-year period in more than 15,000 years of preindustrial simulations exhibits a correlation coefficient as positive as that between the observed and ensemble mean ALL trend patterns, suggesting that external radiative forcing was a driving factor in large-scale cloud changes from the 1980s to the 2000s.

One prominent feature of Fig. 1, and a robust prediction by climate models, is the widespread reduction in cloudiness at mid-latitudes^{2,10,12,18}. Figure 2a, b shows trends in zonal mean total cloud amount during the period 1983–2009 for ISCCP and PATMOS-x, and Fig. 2c shows zonal mean differences between the 2002–2014 CERES

¹Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California, USA. ²Department of Earth Sciences, University of California at Riverside, Riverside, California, USA.

³Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California, USA. ⁴Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado, USA.

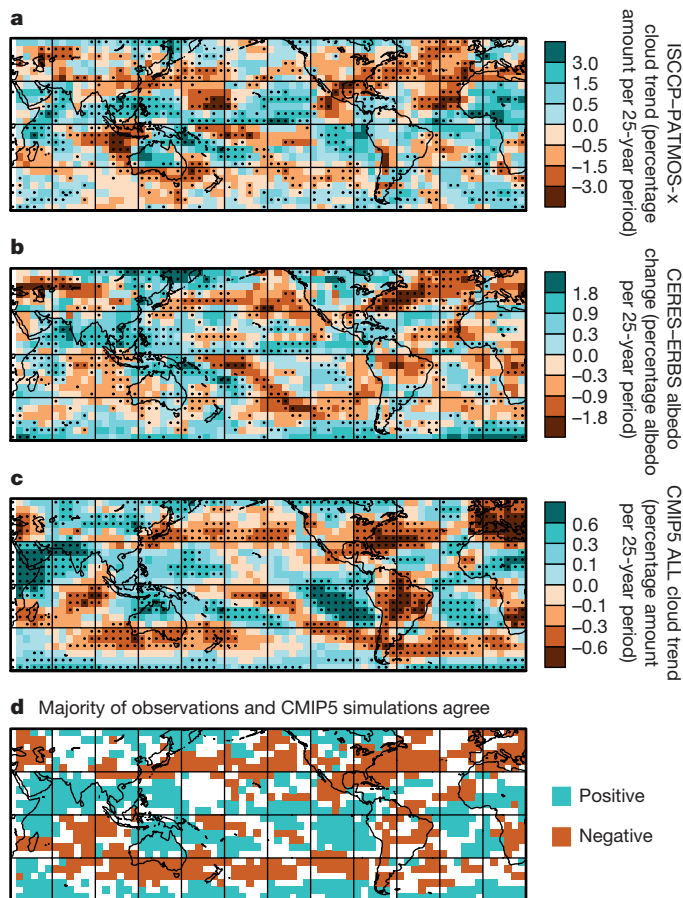


Figure 1 | Change in observed and simulated cloud amount and albedo between the 1980s and 2000s. a, Trend in average of PATMOS-x and ISCCP total cloud amount 1983–2009. **b,** Change in albedo from January 1985–December 1989 (ERBS) to July 2002–June 2014 (CERES). **c,** Trend in ensemble mean total cloud amount 1983–2009 from CMIP5 historical simulations with all radiative forcings (ALL). **d,** Locations where majority of observations and majority of simulations show increases (blue) or decreases (orange). Black dots indicate agreement among all three satellite records on sign of change in **a** and **b** and trend statistical significance ($P < 0.05$ two-sided) in **c**. All trends and changes are relative to the 60°S – 60°N mean change.

albedo and the 1985–1989 ERBS albedo. Every observational record exhibits a decline in cloud amount or albedo at mid-latitudes in both hemispheres that is nearly always statistically significant. The ocean-only MAC-LWP dataset also reports less liquid water path around 40°N and 40°S (Extended Data Fig. 1b). Previous research found evidence for tropical expansion in recent decades¹⁹. Reduced cloudiness around 40°N and 40°S is consistent with a poleward expansion of

Table 1 | Correlation between observed and modelled cloud trend patterns

Spatial pattern	Forcing type				
	ALL	GHG	AA	OZ	NAT
Grid box total cloud amount	0.39 (0.0001) [0.003]	0.21 (0.05) [0.08]	0.00	0.00	0.26 (0.03) [0.04]
Zonal mean total cloud amount	0.80 (0.002) [0.009]	0.62 (0.008) [0.06]	−0.35	0.27	0.69 (0.03) [0.03]
Zonal mean cloud amount in the 50–180 hPa and 180–320 hPa intervals	0.76 (0.003) [0.03]	0.73 (0.004) [0.04]	−0.62		0.73 (0.003) [0.04]

Parentheses and square brackets indicate one-sided P values obtained from the preindustrial simulations shown in Extended Data Fig. 3 and from formal significance tests, respectively.

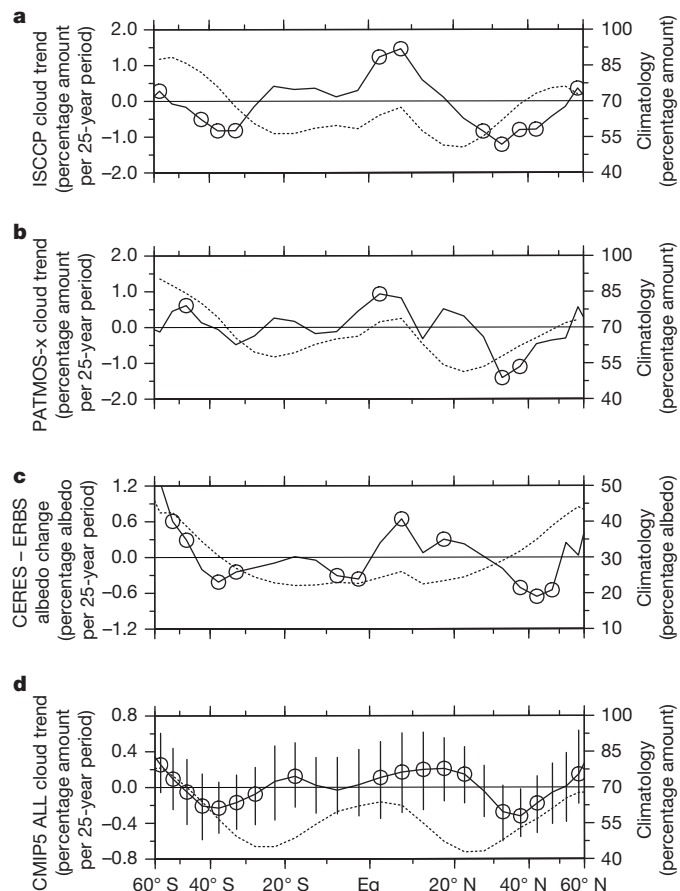


Figure 2 | Zonal mean change in observed and simulated cloud amount and albedo between the 1980s and the 2000s. a, Trend in ISCCP total cloud amount 1983–2009. **b,** Trend in PATMOS-x total cloud amount 1983–2009. **c,** Change in albedo from January 1985–December 1989 (ERBS) to July 2002–June 2014 (CERES). **d,** Trend in ensemble mean total cloud amount 1983–2009 from CMIP5 historical simulations with all radiative forcings (ALL). Zonal mean climatology is dotted, linear trend or change is solid, circles indicate trend statistical significance ($P < 0.05$ two-sided), and bars indicate the interquartile range of individual simulations. All trends and changes are relative to the 60°S – 60°N mean change.

the subtropical dry zone cloud minimum and poleward retreat of the storm-track cloud maximum.

Figure 2d displays trends in zonal mean total cloud amount during the period 1983–2009 from the ALL simulations. Most individual simulations exhibit reduced cloud amount in the mid-latitudes of both hemispheres, and the ensemble mean trends are statistically significant ($P < 0.05$ two-sided). Furthermore, the majority of simulations reproduce the observed increase in cloud amount and albedo occurring in the northern tropics. The spatial correlation between observed and simulated zonal cloud trends is highly significant (Table 1 and Extended Data Fig. 3).

Since the correction procedures applied to the satellite datasets removed any real global mean change that might be present, for maximum comparability we subtracted the 60°S – 60°N average change in total cloud amount from the model output before creating Figs 1c and d, and 2d. Without this adjustment, the ALL ensemble mean cloud amount averaged over 60°S – 60°N decreases by 0.13% over 25 years. Although highly statistically significant ($P < 0.0001$ two-sided), the modelled reduction in 60°S – 60°N average cloud amount during the period 1983–2009 is far smaller than what is detectable by our observational systems. Extended Data Fig. 4a and b shows ALL cloud trends without the subtraction of the 60°S – 60°N average change. They exhibit patterns similar to those seen in Figs 1c and 2d.

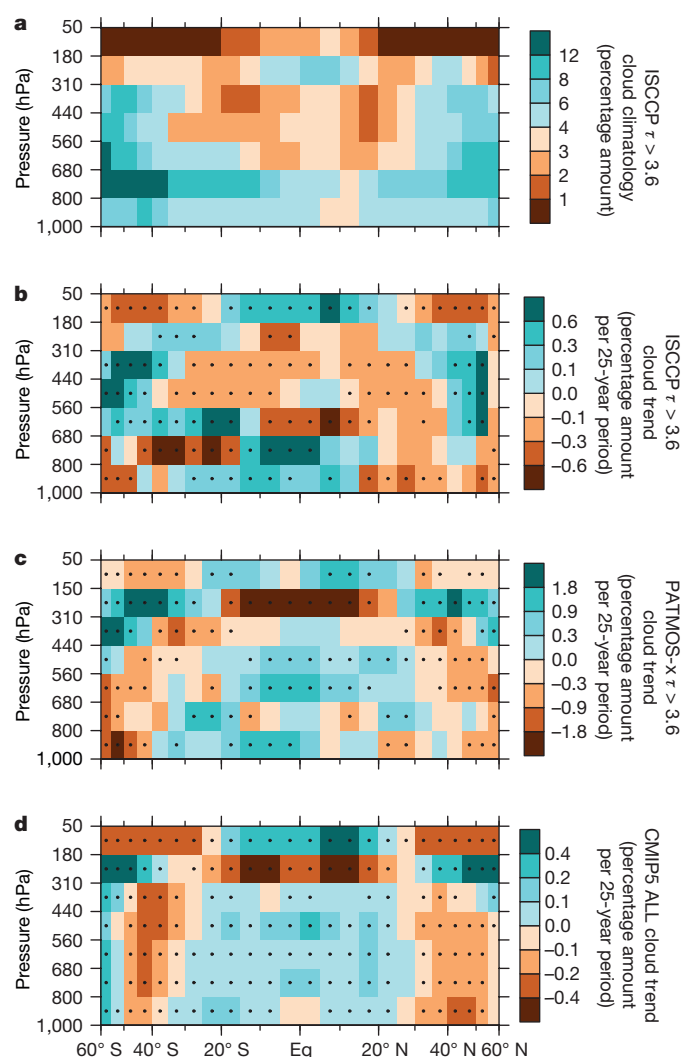


Figure 3 | Zonal mean change in observed and simulated cloud amount during the period 1983–2009 in seven pressure intervals. a, ISCCP climatological cloud amount. **b,** Trend in ISCCP cloud amount 1983–2009. **c,** Trend in PATMOS-x cloud amount 1983–2009. **d,** Trend in ensemble mean cloud amount 1983–2009 from CMIP5 historical simulations with all radiative forcings (ALL). For ISCCP and PATMOS-x, only the amount of clouds with optical thickness $\tau > 3.6$ is plotted. Black dots indicate trend statistical significance ($P < 0.05$ two-sided). All trends are relative to the 60° S– 60° N mean trend for that pressure interval.

Another robust prediction by climate models is the rising height of high cloud tops at all latitudes^{10,12,18,20}. Figure 3a displays ISCCP climatological zonal mean cloud amount within 7 cloud top pressure intervals. Only amounts of clouds with optical thickness greater than 3.6 are plotted to reduce uncertainty in cloud-top pressure retrievals. A local maximum in cloud amount occurs in the 180–310 hPa interval in the tropics, whereas clouds are typically no higher than 310 hPa in the mid-latitude storm tracks, following the latitudinal variation of tropopause height. Figure 3b, c shows that the ISCCP and PATMOS-x zonal mean cloud amounts increased in the 50–180 hPa interval and decreased in the 180–310 hPa interval during the period 1983–2009 in the tropics, consistent with a rise in the tops of the highest clouds. The increase in cloud amount in the 180–310 hPa interval at mid-latitudes similarly suggests a rise in the highest cloud tops.

Figure 3d displays trends in zonal mean cloud amount during the period 1983–2009 from the ALL ensemble mean. The pattern of modelled cloud trends is highly correlated with the satellite record in the 50–180 hPa and 180–310 hPa intervals, suggesting that the observed rise in cloud top is at least partly due to external radiative forcing.

We expect less agreement below these levels because the ISCCP and PATMOS-x satellite retrievals cannot detect lower clouds underneath higher clouds, whereas the models report the exact cloud amount at each level. We note that the negative trends in cloud amount occurring in the 50–180 hPa interval at higher latitudes are relative to the 60° S– 60° N average cloud change for that interval and do not correspond to an actual reduction in cloudiness. Extended Data Fig. 4c, for which the 60° S– 60° N average change was not subtracted, shows that modelled cloud amount in the 50–180 hPa interval merely increases less at higher latitudes than at lower latitudes.

What specific factors are contributing to the observed cloud changes? We addressed this question by examining the additional CMIP5 simulations listed in Extended Data Table 1 with external radiative forcing only from greenhouse gases (GHG), only from anthropogenic aerosol (AA), only from ozone (OZ), and only from natural solar variations and volcanic aerosol (NAT). Extended Data Figs 5, 6 and 7 display the ensemble mean modelled cloud trends for GHG, AA, OZ and NAT simulations. The GHG and NAT simulations both produce modelled cloud trend patterns that are significantly correlated with the observed cloud trend pattern (Table 1 and Extended Data Fig. 3). This includes decreasing total cloud amount at mid-latitudes (GHG and NAT), increasing total cloud amount in the northern tropics (NAT), and increasing cloud amount in the 50–180 hPa interval in the tropics and in the 180–310 hPa interval at mid-latitudes (GHG and NAT). In contrast, the AA and OZ simulations do not produce cloud trends that globally resemble the observed cloud trends, as demonstrated by insufficiently positive correlation (Table 1). The OZ simulations do exhibit reduced cloud amount at Southern Hemisphere mid-latitudes²¹.

Both the GHG and the NAT simulations experience increasing tropospheric temperature and decreasing stratospheric temperature from the 1980s and the 2000s. This is caused by increasing greenhouse gases in the former case and a recovery from the 1982 El Chichón and 1991 Pinatubo volcanic aerosol episodes in the latter case^{22–24}. Tropospheric warming and stratospheric cooling promote an increase in the height of the highest cloud tops^{25,26}, and together with global warming, promote an expansion of the tropical zone and a poleward shift of storm tracks^{27,28}. Depleted stratospheric ozone is an additional factor promoting a poleward shift of the Southern Hemisphere storm track^{21,29}.

The expansion of subtropical dry zones results in less reflection of solar radiation back to space. As cloud tops rise, their greenhouse effect becomes stronger. Both of these cloud changes have a warming effect on climate. Our results suggest that radiative forcing by a combination of anthropogenic greenhouse gases and volcanic aerosol has produced observed cloud changes during the past several decades that exert positive feedbacks on the climate system. We expect that increasing greenhouse gases will cause these cloud trends to continue in the future, unless offset by unpredictable large volcanic eruptions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 November 2015; accepted 7 April 2016.

Published online 11 July 2016.

1. Ramanathan, V. *et al.* Cloud-radiative forcing and climate: results from the Earth Radiation Budget Experiment. *Science* **243**, 57–63 (1989).
2. Boucher, O. *et al.* In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) Ch. 7, 571–657 (Cambridge Univ. Press, 2013).
3. Dufresne, J.-L. & Bony, S. An assessment of the primary sources of spread of global warming estimates from coupled atmosphere–ocean models. *J. Clim.* **21**, 5135–5144 (2008).
4. Evan, A. T., Heidinger, A. K. & Vimont, D. J. Arguments against a physical long-term trend in global ISCCP cloud amounts. *Geophys. Res. Lett.* **34**, L04701 (2007).
5. Norris, J. R. & Evan, A. T. Empirical removal of artifacts from the ISCCP and PATMOS-x satellite cloud records. *J. Atmos. Ocean. Technol.* **32**, 691–702 (2015).
6. Rossow, W. B. & Schiffer, R. A. Advances in understanding clouds from ISCCP. *Bull. Am. Meteorol. Soc.* **80**, 2261–2287 (1999).

7. Heidinger, A. K., Foster, M. J., Walther, A. & Zhao, Z. The Pathfinder Atmospheres Extended (PATMOS-x) AVHRR climate data set. *Bull. Am. Meteorol. Soc.* **95**, 909–922 (2014).
8. Marchand, R. Trends in ISCCP, MISR, and MODIS cloud-top-height and optical-depth histograms. *J. Geophys. Res.* **118**, 1941–1949 (2013).
9. Bender, F. A.-M., Ramanathan, V. & Tselioudis, G. Changes in extratropical storm track cloudiness 1983–2008: observational support for a poleward shift. *Clim. Dyn.* **38**, 2037–2053 (2012).
10. Marvel, K. *et al.* External influences on modeled and observed cloud trends. *J. Clim.* **28**, 4820–4840 (2015).
11. Eastman, R. & Warren, S. G. A 39-yr survey of cloud changes from land stations worldwide 1971–2009: long-term trends, relation to aerosols, and expansion of the Tropical belt. *J. Clim.* **26**, 1286–1303 (2013).
12. Zelinka, M. D., Klein, S. A. & Hartmann, D. L. Computing and partitioning cloud feedbacks using cloud property histograms. Part II: Attribution changes in cloud amount, altitude, and optical depth. *J. Clim.* **25**, 37363754 (2012).
13. Barkstrom, B. E. *et al.* Earth Radiation Budget Experiment (ERBE) archival and April 1985 results. *Bull. Am. Meteorol. Soc.* **70**, 1254–1262 (1989).
14. Wielicki, B. A. *et al.* Clouds and the Earth's Radiant Energy System (CERES): an Earth observing experiment. *Bull. Am. Meteorol. Soc.* **77**, 853–868 (1996).
15. Loeb, N. G. *et al.* Toward optimal closure of the Earth's TOA radiation budget. *J. Clim.* **22**, 748–766 (2009).
16. O'Dell, C. W., Wentz, F. J. & Bennartz, R. Cloud liquid water path from satellite-based passive microwave observations: a new climatology over the global oceans. *J. Clim.* **21**, 1721–1739 (2008).
17. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experimental design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
18. Wetherald, R. T. & Manabe, S. Cloud feedback processes in a general circulation model. *J. Atmos. Sci.* **45**, 1397–1416 (1988).
19. Seidel, D. J., Fu, Q., Randel, W. J. & Reichler, T. J. Widening of the tropical belt in a changing climate. *Nat. Geosci.* **1**, 21–24 (2008).
20. Chepfer, H., Noel, V., Winker, D. & Chiriaco, M. Where and when will we observe cloud changes due to climate warming? *Geophys. Res. Lett.* **41**, 8387–8395 (2014).
21. Grise, K. M., Polvani, L. M., Tselioudis, G., Wu, Y. & Zelinka, M. D. The ozone hole indirect effect: cloud-radiative anomalies accompanying the poleward shift of the eddy-driven jet in the Southern Hemisphere. *Geophys. Res. Lett.* **40**, 3688–3692 (2013).
22. Vallis, G. K., Zurita-Gotor, P., Cairns, C. & Kidston, J. Response of the large-scale structure of the atmosphere to global warming. *Q. J. R. Meteorol. Soc.* **141**, 1479–1501 (2015).
23. Dutton, E. G. & Christy, J. R. Solar radiative forcing at selected locations and evidence for global lower tropospheric cooling following the eruptions of El Chichón and Pinatubo. *Geophys. Res. Lett.* **19**, 2313–2316 (1992).
24. Labitzke, K. & McCormick, M. P. Stratospheric temperature increases due to Pinatubo aerosols. *Geophys. Res. Lett.* **19**, 207–210 (1992).
25. Hartmann, D. L. & Larson, K. An important constraint on tropical cloud-climate feedback. *Geophys. Res. Lett.* **29**, <http://dx.doi.org/10.1029/2002GL015835> (2002).
26. Zelinka, M. D. & Hartmann, D. L. Why is longwave cloud feedback positive? *J. Geophys. Res.* **115**, D16117 (2010).
27. Held, I. M. & Hou, A. Y. Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.* **37**, 515–533 (1980).
28. Lu, J., Vecchi, G. A. & Reichler, Y. Expansion of the Hadley cell under global warming. *Geophys. Res. Lett.* **34**, L06805 (2007).
29. Polvani, L. M., Waugh, D. W., Correa, G. J. P. & Son, S. W. Stratospheric ozone depletion: the main driver of Twentieth-Century atmospheric circulation changes in the Southern Hemisphere. *J. Clim.* **24**, 795–812 (2011).

Acknowledgements National Oceanic and Atmospheric Administration (NOAA) awards NA100AR4310140 and NA100AR4310141 supported work by J.R.N. and A.T.E. The efforts of M.D.Z. and S.A.K. were supported by the US Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research through its Regional and Global Climate Modeling Program and were performed under the auspices of the DOE by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Part of the work by M.D.Z. was supported by the National Aeronautics and Space Administration (NASA) New Investigator Program (NNH14AX831). The MAC-LWP climatology is supported by the NASA MEaSUREs Program (NNH12ZDA001N). We acknowledge the World Climate Research Programme's (WCRP's) Working Group on Coupled Modelling, which is responsible for the Coupled Model Intercomparison Project (CMIP), and we thank the climate modelling groups for producing and making available their model output. For CMIP the US DOE's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author Contributions J.R.N. designed the study, provided ERBS, CERES, and ISCCP data, did the main analysis, and wrote the paper; R.J.A. provided standard model cloud output for CMIP5 simulations and analysed CMIP5 meteorological output; A.T.E. provided corrected PATMOS-x data; M.D.Z. provided CMIP5 COSP cloud output; C.W.O. provided MAC-LWP liquid water path data; and S.A.K. provided background information and ideas. All authors discussed the results and commented on the manuscript.

Author Information Corrected ISCCP and PATMOS-x cloud amount data are available from the Research Data Archive at NCAR at <http://dx.doi.org/10.5065/D62J68XR>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R.N. (jnorrisc@ucsd.edu).

METHODS

Satellite datasets. ISCCP provides values of cloud amount in seven cloud top pressure intervals and six optical thickness intervals (that is, cloud amount for each of 42 'cloud types') from July 1983 to December 2009 (ref. 6). Total cloud amount is the sum over all intervals. Cloud-top pressure is most accurately identified when clouds are nearly opaque at thermal infrared wavelengths. This occurs when cloud optical thickness at visible wavelengths is greater than 3.6. Geostationary satellites are the primary contributors to the ISCCP cloud record. We downloaded ISCCP D1 data from the Atmospheric Science Data Center located at NASA Langley Research Center and applied a correction procedure to remove spurious variability associated with changes in satellite orbits, satellite calibration and ancillary data⁵. We note that the correction procedure removes any real global mean cloud variability, so all trends presented in this study are with respect to an unknown global mean trend, which could be zero. The present study uses only daytime observations (defined as solar zenith angle $<78^\circ$) because visible radiances are required to retrieve cloud optical thickness. We found that trends in total cloud amount retrieved from day and night infrared radiances are very similar to trends in total cloud amount retrieved from daytime visual+infrared radiances.

PATMOS-x provides cloud amount in seven cloud-top pressure intervals and six optical thickness intervals starting in October 1981 (ref. 7). The present study uses data from January 1983 to December 2009 for consistency with ISCCP. The total cloud amount is the sum over all intervals. Polar-orbiting satellites are the only contributors to the PATMOS-x cloud record. We downloaded PATMOS-x Version 5 Level 3 'GEWEX' data from <https://cimss.ssec.wisc.edu/patmosx/access.html> and applied a correction procedure to remove spurious variability associated with changes in satellite orbits, satellite calibration, and ancillary data⁵. As for ISCCP, the correction procedure removes any real global mean cloud variability. For consistency over the entire PATMOS-x record, we use products retrieved only from the daytime pass of the 'afternoon' satellites.

The passive remote sensing techniques employed by ISCCP and PATMOS-x can have difficulty identifying the occurrence of optically thin cirrus overlying optically thick lower cloud and can underestimate the height of the cloud top when cloud particle density is sparse in the upper portion of an optically thick cloud^{30,31}. ISCCP suffers more from remote sensing limitations than PATMOS-x since the latter dataset uses more wavelengths and thus has more information available to retrieve cloud properties⁷. The result is a downward bias in reported cloud-top height relative to that obtained from active remote-sensing techniques, for which only a short record is available. ISCCP correspondingly underestimates the cloud amount in the 50–180 hPa and 180–310 hPa pressure intervals compared to active remote sensing. Despite the bias, a real increase over time in the height of the highest cloud tops will nonetheless be reported by ISCCP as an increase in the amount of cloudiness in the higher elevation pressure interval and a decrease in the lower elevation pressure intervals. The bias may produce an underestimate in the magnitude of cloud trends since ISCCP climatological cloud amount is underestimated, but this does not undermine our analysis because we compare the relative spatial patterns of observed and modelled cloud change rather than the absolute magnitudes of observed and modelled cloud change.

Albedo is a useful parameter for our investigation because variability in cloud amount is by far the dominant contributor to variability in albedo outside ice-covered regions. Variability in cloud optical thickness is a secondary contributor. ERBS albedo values are available for November 1984 through February 1990, but we use the January 1985–December 1989 climatology for simplicity¹³. We also use measurements from ERBS only because the other two satellites contributing to the Earth Radiation Budget Experiment (NOAA9 and NOAA10) were not available for the entire period. We note that ERBS was in a precessing orbit and sampled the entire diurnal cycle. CERES Energy Balanced And Filled (EBAF) Ed2.8 reflected solar radiation values are available from the morning satellite Terra starting in March 2000 and from the afternoon satellite Aqua starting in July 2002 (ref. 14). Since Terra and Aqua are in Sun-synchronous orbits CERES EBAF uses geostationary satellites to fill out the diurnal cycle of cloudiness not sampled by Terra and Aqua¹⁵. To ensure consistent sampling of the diurnal cycle and seasonal cycle, we constructed a climatology for July 2002–June 2014. We then divided reflected solar radiation by incoming solar radiation at the top of the atmosphere to calculate albedo. ERBS data were obtained from a CD-ROM provided by the Atmospheric Science Data Center located at NASA Langley Research Center, and CERES data were obtained from the NASA Langley Research Center CERES ordering tool at (<http://ceres.larc.nasa.gov/>). Although the datasets are individually well calibrated, there is no absolute calibration between ERBS and CERES. To bring them to a common reference point, we multiplied the ERBS albedo by a constant factor so that ERBS and CERES had the same climatological annual albedo averaged over 60°S – 60°N . This means that CERES–ERBS differences are relative to an unknown global mean difference that could be zero.

Version 4 of the MAC-LWP dataset for January 1988–December 2014 provides a useful complement to measurements of cloud amount and albedo¹⁶. The MAC-LWP dataset synthesizes passive microwave retrievals from 12 different sensors using the Remote Sensing Systems version 7 ocean algorithm³². Liquid water path is the spatially averaged vertically integrated amount of cloud liquid water within a satellite footprint. Cloud-free areas contribute a value of zero to the spatial average within the footprint. Liquid water path increases as clouds become more horizontally extensive (that is, as cloud amount increases). It also increases as clouds become vertically thicker and as cloud water concentration becomes larger. The dataset does not include contributions from cloud ice, and retrievals are available only over open ocean. To provide a similar basis for comparison to the ISCCP, PATMOS-x, and ERBS/CERES datasets, from which global mean variability was removed in the correction and adjustment process, we subtracted the 60°S – 60°N average liquid water path from the value at each grid box for each month. This has little impact on the spatial distribution of trends.

CMIP5 simulations. The CMIP5 multi-model dataset provides a large number of global climate model simulations for various forcing scenarios¹⁷. The historical simulations span ~ 1850 –2005 and include time-varying radiative forcings such as greenhouse gases, ozone, anthropogenic and volcanic aerosols, solar output, and land use changes (ALL). We extended the CMIP5 ALL simulations beyond their nominal ending year of 2005 by adding follow-on years through to 2009 with radiative forcing from representative concentration pathway 4.5 (or if not available, the historical extended experiment or representative concentration pathway 8.5). Total cloud amount is available from 107 realizations from 33 models, and cloud amount in each vertical layer is available from 76 realizations from 27 models (Extended Data Table 1). We calculated the ensemble mean as a simple average of all available realizations. Some models provided only one realization and other models provided up to ten realizations for the same external forcing. Natural internal variability across the simulations tends to cancel in the ensemble mean, leaving behind the radiatively forced component of cloud change. The ensemble mean has smaller trend amplitudes than any one realization or the observations due to this reduction of natural internal variability.

A smaller set of CMIP5 models provided additional simulations with external radiative forcing only from anthropogenic greenhouse gases (GHG), only from anthropogenic aerosol (AA), only from ozone (OZ), and only from natural solar variability and volcanic aerosol (NAT) (Extended Data Table 1). A few models included ozone variability in GHG simulations, but we excluded these from our analysis to avoid confusion about forcing factors. Total cloud amount is available from 44 realizations from 14 models for GHG, from 33 realizations from 11 models for AA, from 11 realizations from 3 models for OZ, and from 37 realizations from 10 models for NAT. Cloud amount in each vertical layer is available from 35 realizations from 12 models for GHG, 32 realizations from 11 models for AA, from 1 realization from 1 model for OZ, and 28 realizations from 8 models for NAT. We did not analyse cloud amount in vertical layers for OZ since only one realization was available.

For many of these models, GHG, AA, OZ, and NAT output was available only until 2005. To maximize the number of realizations, we chose the 1979–2005 interval to calculate trends for GHG and AA since this time period has the same length as the ISCCP and PATMOS-x records. The four-year shift for trend calculation should not matter for the GHG and AA simulations since greenhouse gas and aerosol emissions vary on multidecadal rather than interannual timescales. Timing matters more for OZ and NAT because the former includes stabilization of the ozone hole in the 2000s and the latter includes volcanic eruptions, so we chose only those models providing output for the full 1983–2009 time period. We note that the set of contributing models and numbers of realizations is not identical for the ALL, GHG, AA, OZ, and NAT simulations. We chose to use all available simulations from each forcing scenario because restricting our comparison to only those models and numbers of realizations in common would result in a much smaller sample size.

Most CMIP5 models provided multicentury simulations of preindustrial conditions with no anthropogenic or natural external radiative forcing as a control case (Extended Data Table 2). Cloud variability in these simulations results only from natural internal variability of the coupled ocean–atmosphere–land system, including El Niño/Southern Oscillation (ENSO). Total cloud amount is available for a total of 15,807 years from 27 models, and cloud amount in each vertical layer is available for a total of 7,311 years from 13 models.

Output from CMIP5 simulations was downloaded from the Earth System Grid Federation. To provide a similar basis for comparison to the ISCCP, PATMOS-x, and ERBS/CERES datasets, from which global mean variability was removed in the correction and adjustment process, we subtracted the 60°S – 60°N ocean average cloud amount from the value at each ocean grid box for each month and subtracted the 60°S – 60°N land average cloud amount from the value at each land grid box.

With the exception of cloud amount in the 50–180 hPa interval, this has little impact on the spatial distribution of trends. For zonally averaged cloud amount in each vertical layer, we subtracted the 60° S–60° N average.

Since the CMIP5 models do not routinely report amount of cloudiness in various optical thickness intervals, we could not limit our analysis of modelled clouds in Fig. 3 to only those with optical thickness greater than 3.6. Cloud amount from the satellite records also differs from standard cloud amount output from CMIP5 models in that the former do not detect clouds with optical thickness less than about 0.3, whereas the latter report the amount of all clouds, even very optically thin ones. Another difference is that CMIP5 models report actual cloud amount at each model layer, not cloud amount unobscured by higher clouds as do the ISCCP and PATMOS-x satellite datasets. For a closer comparison to observations, a few CMIP5 models incorporated the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (COSP)³³. This software produced model cloud output according to how it would be detected through the limitations of satellite retrieval, most often ISCCP. Since it is computationally expensive, COSP cloud output is available only for 13 ALL realizations from seven models (Extended Data Table 1). We use standard model cloud output in order to have a larger sample size but obtain similar but noisier results if we use COSP cloud output from an ISCCP satellite simulator (Extended Data Fig. 8).

Data analysis. ISCCP, PATMOS-x, and ERBS data on a $2.5^\circ \times 2.5^\circ$ latitude-longitude grid and CERES and MAC-LWP data on a $1^\circ \times 1^\circ$ grid were spatially averaged to a common $5^\circ \times 5^\circ$ grid. Output from CMIP5 models on a variety of grid sizes were bilinearly interpolated to a $5^\circ \times 5^\circ$ grid. We linearly interpolated cloud amount from model layers to a common vertical grid with regular 50 hPa spacing in pressure. To display modelled cloud trends in the figures, we linearly interpolated trends from the 50 hPa regular grid to the midpoints of the seven pressure intervals used by the satellite datasets. All calculations are performed on anomalies from the long-term mean. In all cases of spatial averaging and spatial correlation, grid box values were weighted by the cosine of the grid box centre latitude to account for the variation of grid box area with latitude. We restrict our analysis to latitudes equatorward of 60° because passive retrieval of cloud properties by satellite is difficult over bright and cold surfaces, and no visible retrievals can be made during polar night.

We use least-squares linear trends or the average difference between two time periods as convenient means of summarizing change over time. Two-sided *P* values for the trends are determined using a conventional Student's *t*-test with an effective sample size that takes temporal autocorrelation into account.

For simplicity of comparison with modelled total cloud amount trends in calculating correlations in Table 1, we averaged cloud/albedo changes from all cloud amount and albedo datasets together before comparing to the ensemble mean total cloud amount trends from the CMIP5 ALL, GHG, AA, OZ, NAT and preindustrial simulations. Since cloud amount and albedo have different physical units, we standardized the grid box changes before averaging. Specifically, we divided each ISCCP grid box cloud amount trend by the standard deviation of all ISCCP grid box cloud amount trends, each PATMOS-x grid box cloud trend by the standard deviation of all PATMOS-x grid box cloud amount trends, and each grid box albedo change by the standard deviation of all grid box albedo changes. For simplicity of comparison with modelled cloud amount trends in seven pressure intervals, we averaged cloud trends from ISCCP and PATMOS-x datasets together.

Our null hypothesis is that the observed cloud changes result purely from natural internal variability. If so, there should be no systematic relationship between the spatial pattern of cloud trends generated by natural internal variability and the spatial pattern of cloud trends generated by external radiative forcing. The former is represented by individual preindustrial simulations, each with different realizations of natural internal variability, and the latter is represented by the ensemble mean of simulations with external radiative forcing, in which natural variability has been largely averaged out. The suitability of this null hypothesis can be demonstrated by calculating the distribution of spatial correlation coefficients (Pearson's *r*) between the pattern of cloud trends from the ensemble mean of forced simulations (ALL, GHG, or NAT) and the pattern of cloud trends from time periods of similar length obtained from preindustrial control simulations. We build up each null distribution by calculating cloud trends and spatial correlation values during a rolling 27-year period (that is, years 1–27, years 2–28, years 3–29, and so on) through the preindustrial control simulation for each model. Extended Data Fig. 3 displays the

frequency distributions of the calculated correlation values. There are 15,104 time periods for total cloud amount and 6,973 time periods for cloud amount in vertical layers. The mean and median correlation values of the null distributions are zero, as expected, and the total area under each frequency distribution is equal to unity.

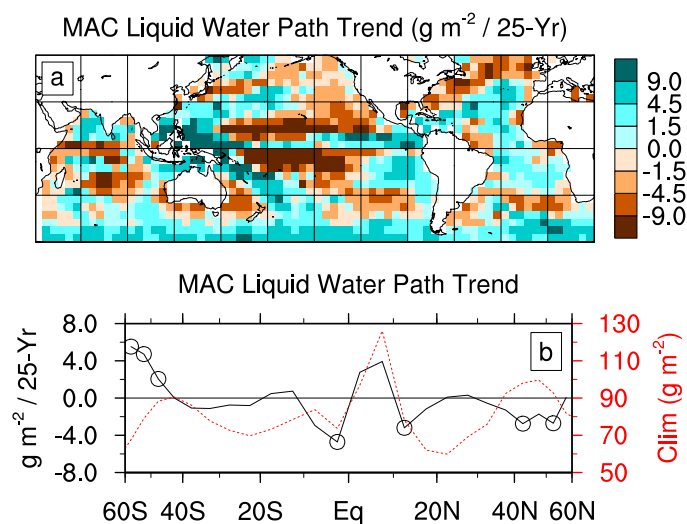
Our alternative hypothesis is that external radiative forcing was a contributing factor in producing the observed cloud trends. If so, we expect a positive spatial correlation between the observed trend pattern and the trend pattern from the ensemble mean of simulations with external radiative forcing (values shown as vertical lines in Extended Data Fig. 3). The *P* value for a particular spatial correlation value *r* is simply the fraction of correlation values from the preindustrial control simulations with values more positive than *r* (that is, the fraction of area under the frequency distribution to the right of the vertical line). For simplicity, we calculate *P* values with respect to a null distribution built from spatial correlation values for cloud trends from single time periods. Another option is to build a null distribution from spatial correlation values for cloud trends from ensemble means of multiple, randomly-selected time periods. Our results are the same for either approach.

For corroboration of *P* values calculated with respect to cloud trend patterns from the preindustrial control simulations, we additionally computed one-sided *P* values using a conventional Student's *t*-test for statistical significance of Pearson's *r*. In this case, a critical parameter is the effective number of spatial degrees of freedom³⁴. We determined that there are 51 spatial degrees of freedom between 60° S and 60° N for the observed total cloud amount in grid boxes. Simplistically considered, this corresponds to a set of boxes slightly larger than those outlined by the latitude-longitude grid lines in Fig. 1, if apportioned equally. However, remote teleconnections contribute to reduced spatial degrees of freedom in addition to local spatial coherence. Zonal means have substantially fewer degrees of freedom (8 for total cloud amount and 7 for cloud amount in the 50–180 hPa and 180–310 hPa pressure intervals). This corresponds to about 15° spacing in latitude. *P* values obtained from formal tests are in some cases substantially larger than those obtained from the preindustrial control simulations (Table 1 and Extended Data Fig. 3), suggesting that the actual number of effective spatial degrees of freedom may be larger than that indicated by the method we used³⁴.

ENSO-like variability. The dominant source of multiyear natural variability in the climate system is the El Niño/Southern Oscillation (ENSO) phenomenon. Variability occurring at interdecadal time scales, especially over the Pacific basin, exhibits a pattern similar to that of ENSO³⁵. We investigated whether the observed cloud changes are a manifestation of ENSO-like variability by calculating the correlation of the spatial pattern of cloud trends with the spatial pattern of the difference between observed La Niña composite cloud anomalies and El Niño composite cloud anomalies (figure not shown due to space limitations). The correlation between the observed La Niña–El Niño pattern and the observed trend pattern is only 0.13, and the spatial correlation between the observed La Niña–El Niño pattern and the ensemble mean ALL trend pattern is only 0.14. Considering that the spatial correlation between the observed trend pattern and the ensemble mean ALL trend pattern is 0.39 (Table 1), we think ENSO-like variability cannot be a major contributor to the global pattern of cloud change between the 1980s and the 2000s.

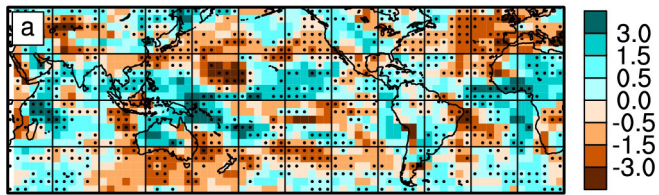
Code availability. All plots and calculations were produced with custom code using the NCAR Command Language (NCL) Version 6.3.0 (<http://dx.doi.org/10.5065/D6WD3XH5>). This code can be obtained by contacting the corresponding author.

30. Mace, G. G. & Wrenn, F. J. Evaluation of the hydrometeor layers in the East and West Pacific within ISCCP cloud-top pressure-optical depth bins using merged CloudSat and CALIPSO data. *J. Clim.* **26**, 9429–9444 (2013).
31. Sherwood, S. C., Chae, J.-H., Minnis, P. & McGill, M. M., Underestimation of deep convective cloud tops by thermal imagery. *Geophys. Res. Lett.* **31**, L11102 (2004).
32. Hilburn, K. A. & Wentz, F. J. Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm (UMORA). *J. Appl. Meteorol. Climatol.* **47**, 778–794 (2008).
33. Bodas-Salcedo, A. et al. 2011: COSP: Satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.* **92**, 1023–1043 (2011).
34. Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. & Bladé, I. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12**, 1990–2009 (1999).
35. Zhang, Y., Wallace, J. M. & Battisti, D. S. ENSO-like interdecadal variability: 1900–93. *J. Clim.* **10**, 1004–1020 (1997).

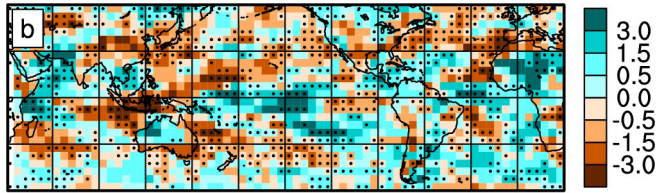


Extended Data Figure 1 | Change in observed liquid water path during the period 1988–2014. **a**, Linear trend in MAC-LWP liquid water path during January 1988 to December 2014. **b**, Zonal mean climatology (red) and trend (black) in MAC-LWP liquid water path during January 1988–December 2014 given as g m^{-2} per 25-year period. Circles indicate trend statistical significance ($P < 0.05$ two-sided). All trends are relative to the 60°S – 60°N mean trend given as g m^{-2} per 25 year period.

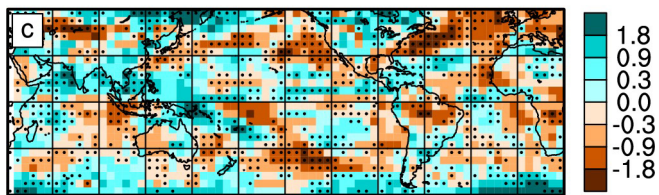
ISCCP Cloud Change (%-Amt / 25-Yr)



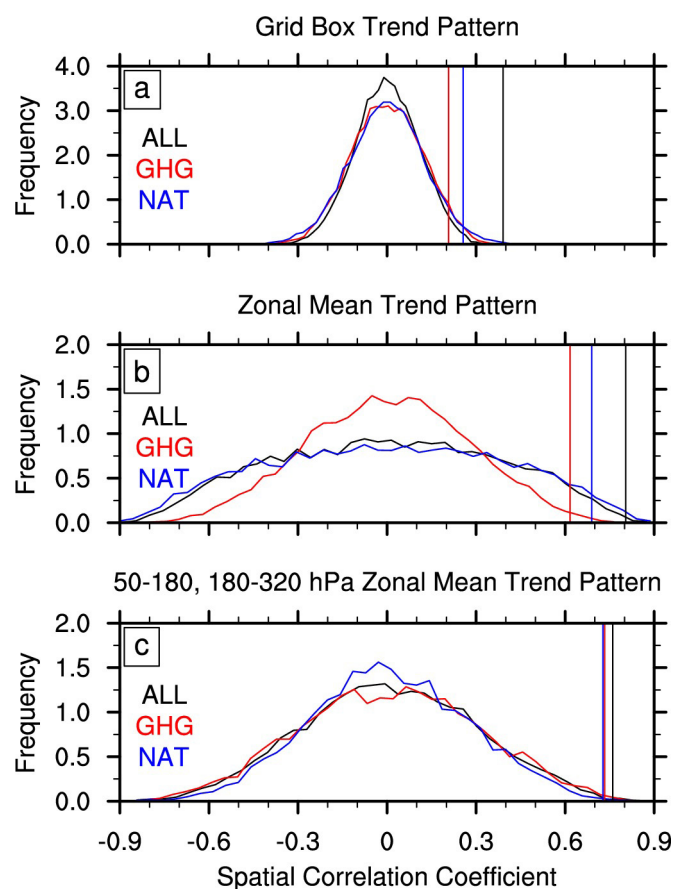
PATMOS-x Cloud Change (%-Amt / 25-Yr)



CERES - ERBS Albedo Change (%-Alb / 25-Yr)

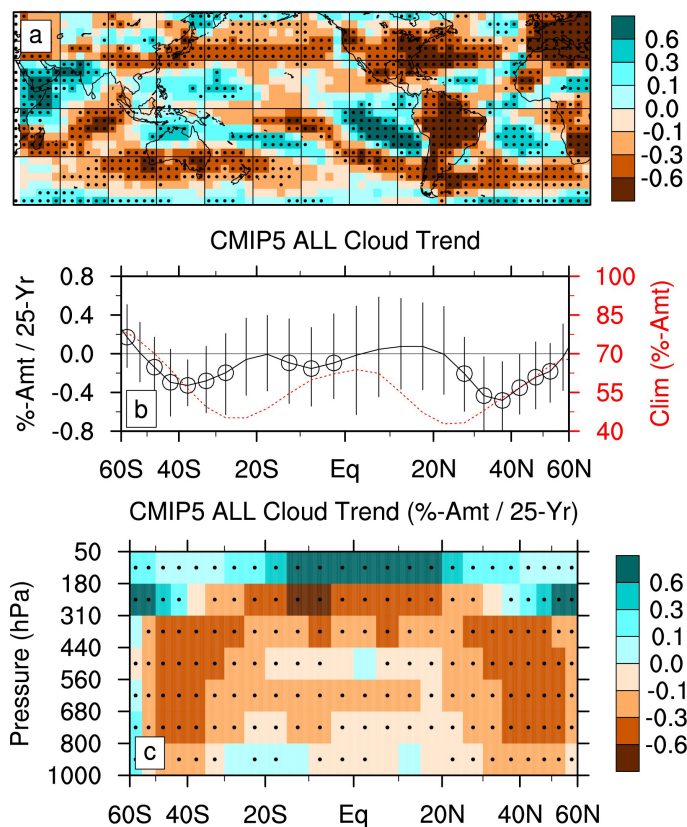


Extended Data Figure 2 | Change in observed cloud amount and albedo between January 1985 to December 1989 and January 2003 to December 2009. **a**, Change in ISCCP total cloud amount given as a percentage amount per 25-year period. **b**, Change in PATMOS-x total cloud amount given as a percentage amount per 25-year period. **c**, Change in ERBS/CERES albedo given as a percentage albedo per 25-year period. Black dots indicate agreement among all three satellite records on sign of change in **a**, **b** and **c**. All changes are relative to the 60° S–60° N mean change.

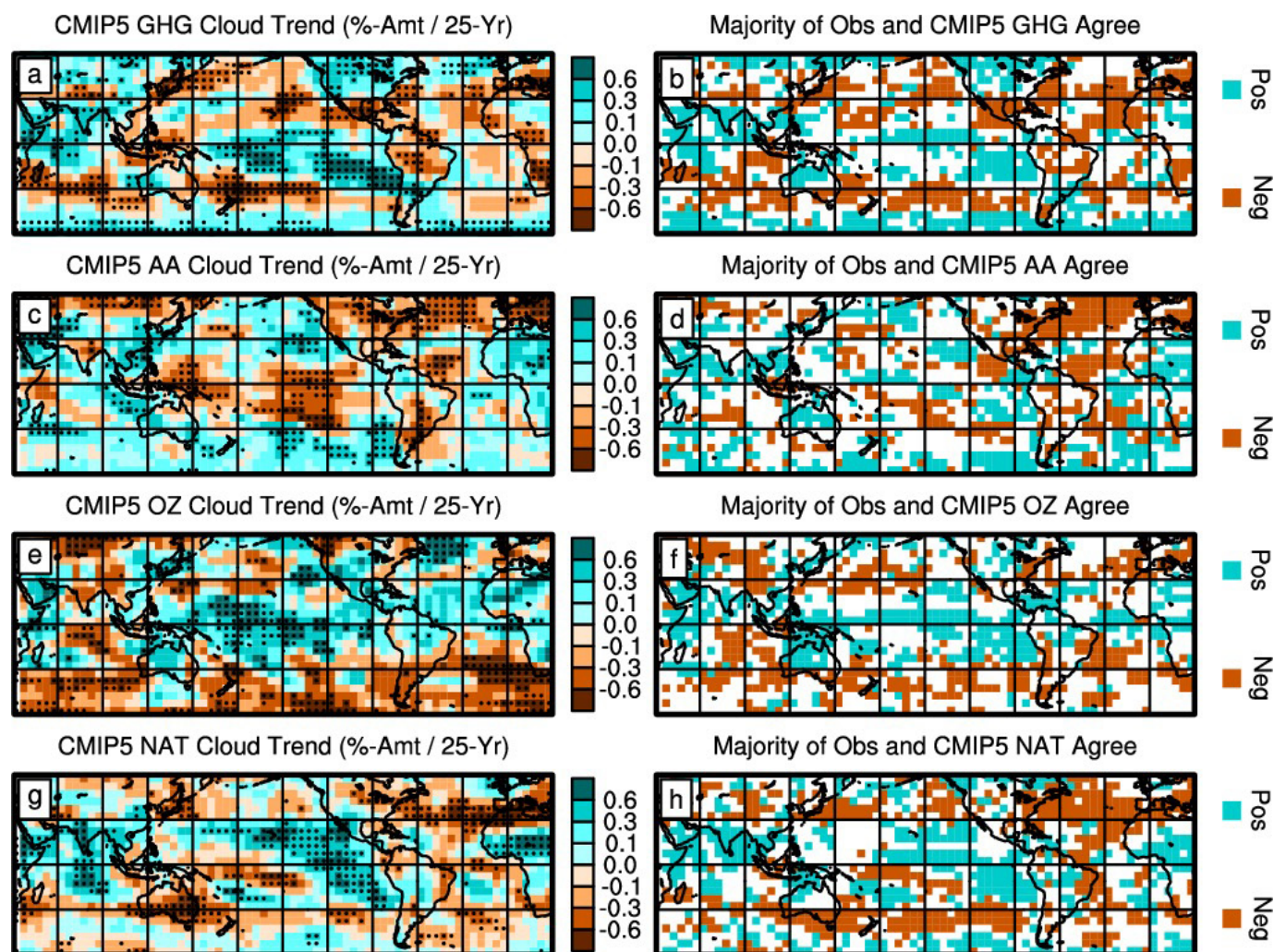


Extended Data Figure 3 | Correlation between forced simulated, unforced simulated, and observed cloud trend patterns. Distribution of correlation between patterns from multiple unforced CMIP5 preindustrial simulations and the ensemble mean of CMIP5 simulations with all radiative forcings (ALL, black), only greenhouse radiative forcings (GHG, red), and only natural radiative forcings (NAT, blue) for 27-year trends. **a**, Grid box total cloud amount. **b**, Zonal mean total cloud amount. **c**, Zonal mean cloud amount in the 50–180 hPa and 180–320 hPa intervals. Vertical lines indicate correlation between the observed pattern and ensemble mean ALL, GHG, or NAT patterns. Fractional area under each distribution to the right of the vertical line is the corresponding *P* value in Table 1.

CMIP5 ALL Cloud Trend (%-Amt / 25-Yr)

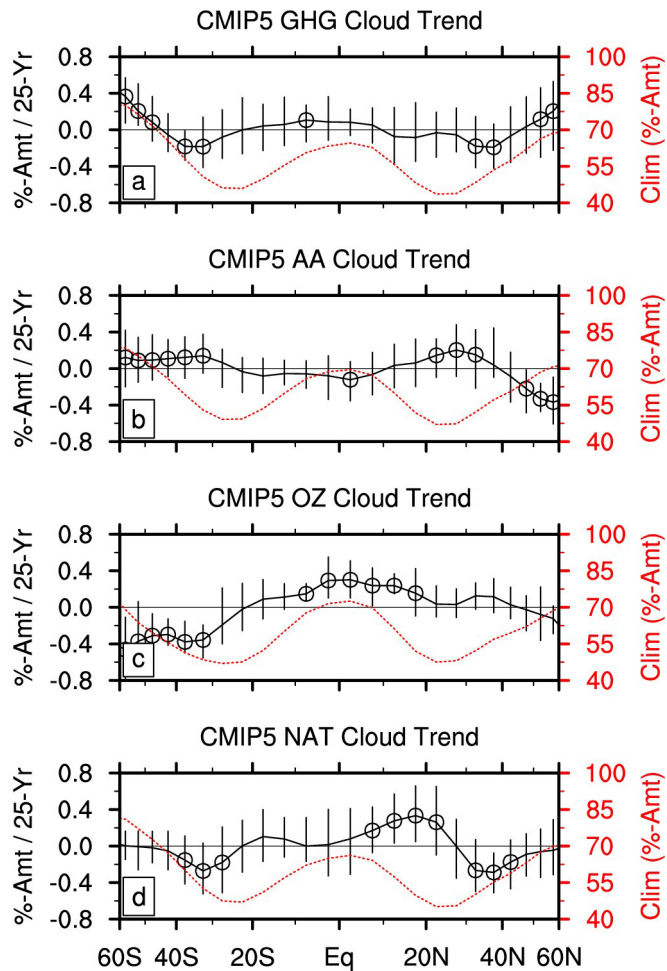


Extended Data Figure 4 | Absolute change in simulated cloud amount during the period 1983–2009. **a**, Linear trend in ensemble mean total cloud amount given as a percentage amount per 25-year period 1983–2009 from CMIP5 historical simulations with all radiative forcings (ALL). Data given as a percentage amount per 25-year period. **b**, Zonal mean climatology (red) and trend (black) in ensemble mean total cloud amount 1983–2009 from ALL simulations. **c**, Zonal mean trend in ensemble mean cloud amount 1983–2009 in seven pressure intervals from ALL simulations. Data given as a percentage amount per 25-year period. Black dots and circles indicate trend statistical significance ($P < 0.05$ two-sided), and bars indicate interquartile range of individual simulations. Unlike Figs 1c, 2d and 3d, trends are not relative to the 60° S–60° N mean trend.

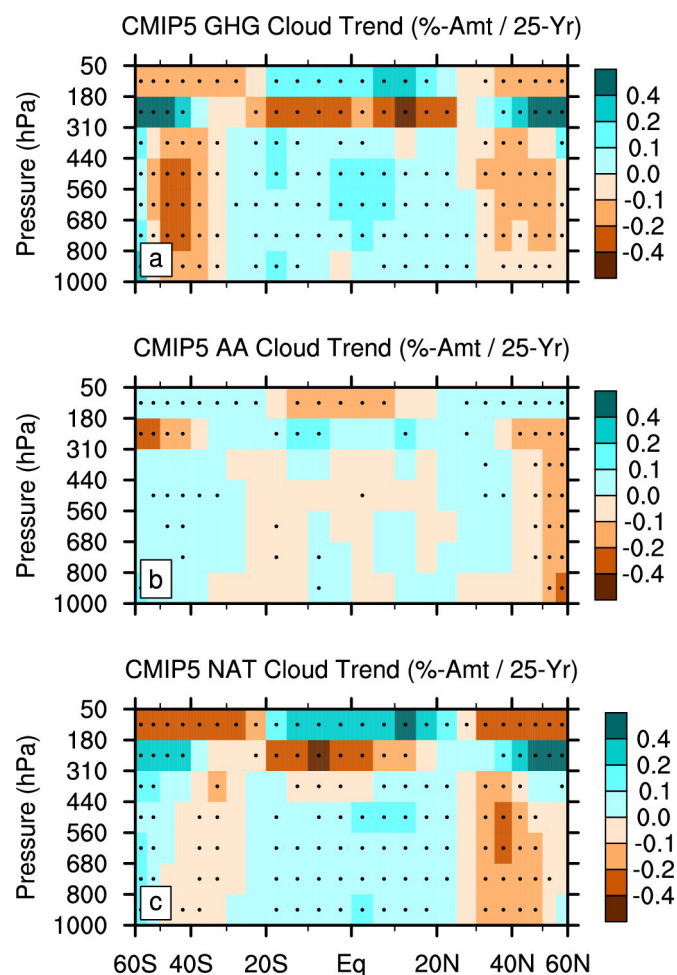


Extended Data Figure 5 | Change in simulated cloud amount between the 1980s and 2000s for different types of forcing. Linear trend in ensemble mean total cloud amount given as a percentage amount per 25-year period. From CMIP5 simulations and locations where the majority of observational (Obs) datasets and the majority of simulations show increasing (blue) or decreasing (orange) cloud amount or albedo.

a, b, Only greenhouse gas (GHG) forcing 1979–2005. c, d, Only anthropogenic aerosol (AA) forcing 1979–2005. e, f, Only ozone (OZ) forcing 1983–2009. g, h, Only natural (NAT) forcing 1983–2009. Black dots indicate trend statistical significance ($P < 0.05$ two-sided). All trends are relative to the 60° S–60° N mean trend.

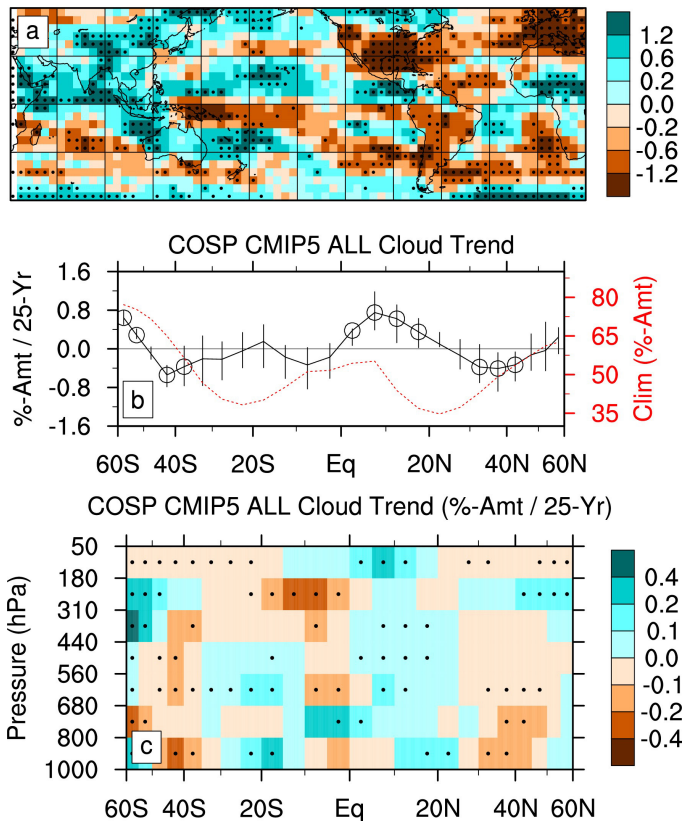


Extended Data Figure 6 | Zonal mean change in simulated cloud amount between the 1980s and 2000s for different types of forcing. Climatology is red and linear trend is black for ensemble mean total cloud amount from CMIP5 simulations. Data given as a percentage amount per 25-year period. **a**, Only greenhouse gas (GHG) forcing 1979–2005. **b**, Only anthropogenic aerosol (AA) forcing 1979–2005. **c**, Only ozone (OZ) forcing 1983–2009. **d**, Only natural (NAT) forcing 1983–2009. Circles indicate trend statistical significance ($P < 0.05$ two-sided) and bars indicate interquartile range of individual simulations. All trends are relative to the 60° S–60° N mean trend.



Extended Data Figure 7 | Zonal mean change in simulated cloud amount between the 1980s and 2000s in seven pressure intervals for different types of forcing. Linear trend for ensemble mean cloud amount from CMIP5 simulations. **a**, Only greenhouse gas (GHG) forcing 1979–2005. Data given as a percentage amount per 25-year period. **b**, Only anthropogenic aerosol (AA) forcing 1979–2005. **c**, Only natural (NAT) forcing 1983–2009. Black dots indicate trend statistical significance ($P < 0.05$ two-sided). All trends are relative to the 60° S–60° N mean trend for that pressure interval.

COSP CMIP5 ALL Cloud Trend (%-Amt / 25-Yr)



Extended Data Figure 8 | Change in simulated cloud amount during the period 1983–2009 from COSP. **a**, Linear trend in ensemble mean total cloud amount 1983–2009 from CMIP5 historical simulations with all radiative forcings (ALL). Data given as a percentage amount per 25-year period. **b**, Zonal mean climatology (red) and trend (black) in ensemble mean total cloud amount 1983–2009 from ALL simulations. **c**, Zonal mean trend in ensemble mean cloud amount 1983–2009 in seven pressure intervals from ALL simulations. Black dots and circles indicate trend statistical significance ($P < 0.05$ two-sided), and bars indicate interquartile range of individual simulations. All trends relative to the 60° S–60° N mean trend.

Extended Data Table 1 | CMIP5 models and number of simulations used for each forcing experiment

Model	ALL		GHG	AA	OZ	NAT
	Standard	COSP				
ACCESS1-0	(1)					
ACCESS1-3	(1)					
BCC-CSM1-1	3		1			1
BCC-CSM1-1-m	3					
BNU-ESM	1		1			
CanESM2	5	1	5	5		5
CCSM4	6		3	3		
CESM1-CAM5	3					
CESM1-CAM5-1-FV2	1		2	2		
CESM1-WACCM	3					
CMCC-CM	1					
CMCC-CMS	1					
CNRM-CM5	(10)		(5)			(5)
CSIRO-Mk3-6-0	10		5	5		5
EC-EARTH	(5)					
FGOALS-g2			1	2	1	
GFDL-CM3	1			3		
GFDL-ESM2G	1					
GFDL-ESM2M	1		1	1		
GISS-E2-H	5		5	5	(5)	5
GISS-E2-H-CC	1					
GISS-E2-R	5		5	5	(5)	5
GISS-E2-R-CC	1					
HadCM3	(10)					
HadGEM2-ES	(4)	1	(4)			(4)
IPSL-CM5A-LR	4		3	1		3
IPSL-CM5A-MR	1		3			3
MIROC5	5	5				
MIROC-ESM	1	3				
MIROC-ESM-CHEM	1	1				
MPI-ESM-LR	3	1				
MPI-ESM-MR	3					
MRI-CGCM3	3	1				
NorESM1-M	3			1		1

Parentheses indicate unavailability of vertical distribution of cloud amount. Empty cells indicate unavailability of model output for that experiment.

Extended Data Table 2 | CMIP5 models and number of years in each preindustrial control simulation

Model	Total Cloud Amount	Layer Cloud Amount
ACCESS1-0	500	
ACCESS1-3	500	
BCC-CSM1-1	500	500
BCC-CSM1-1-m	400	400
BNU-ESM	559	559
CanESM2	996	796
CCSM4	1000	1000
CESM1-CAM5	319	
CESM1-WACCM	200	
CNRM-CM5	850	
CSIRO-Mk3-6-0	500	
FGOALS-g2	700	
GFDL-CM3	500	500
GFDL-ESM2G	500	500
GFDL-ESM2M	500	500
GISS-E2-H	540	
GISS-E2-R	550	
HadGEM2-ES	337	
IPSL-CM5A-LR	1000	
IPSL-CM5A-MR	300	
MIROC5	670	670
MIROC-ESM	630	630
MIROC-ESM-CHEM	255	255
MPI-ESM-LR	1000	
MPI-ESM-MR	1000	
MRI-CGCM3	500	500
NorESM1-M	501	501

Empty cells indicate unavailability of model output for that experiment.

A novel excitatory network for the control of breathing

Tatiana M. Anderson^{1,2*}, Alfredo J. Garcia III^{1*}, Nathan A. Baertsch¹, Julia Pollak¹, Jacob C. Bloom¹, Aguan D. Wei¹, Karan G. Rai¹ & Jan-Marino Ramirez^{1,3}

Breathing must be tightly coordinated with other behaviours such as vocalization, swallowing, and coughing. These behaviours occur after inspiration, during a respiratory phase termed postinspiration¹. Failure to coordinate postinspiration with inspiration can result in aspiration pneumonia, the leading cause of death in Alzheimer's disease, Parkinson's disease, dementia, and other neurodegenerative diseases². Here we describe an excitatory network that generates the neuronal correlate of postinspiratory activity in mice. Glutamatergic–cholinergic neurons form the basis of this network, and GABA (γ -aminobutyric acid)-mediated inhibition establishes the timing and coordination relative to inspiration. We refer to this network as the postinspiratory complex (PiCo). The PiCo has autonomous rhythm-generating properties and is necessary and sufficient for postinspiratory activity *in vivo*. The PiCo also shows distinct responses to neuromodulators when compared to other excitatory brainstem networks. On the basis of the discovery of the PiCo, we propose that each of the three phases of breathing is generated by a distinct excitatory network: the pre-Bötzinger complex, which has been linked to inspiration^{3,4}; the PiCo, as described here for the neuronal control of postinspiration; and the lateral parafacial region (pFR), which has been associated with active expiration, a respiratory phase that is recruited during high metabolic demand^{4,5}.

Neurons that are active in phase with postinspiratory activity have previously been identified in the Bötzinger complex (BötC), a region that is primarily comprised of inhibitory neurons^{6,7}. However, the source of excitation that drives this inhibitory network is not well-defined. In this study, we have identified the location and neurochemical phenotype of an excitatory and rhythmogenic neuronal population that is specifically active during postinspiration.

We developed a horizontal slice preparation in postnatal day (P)5–10 mice that captures the ventral extent of the medulla, including the ventral respiratory column⁸ (VRC; Fig. 1a and Extended Data Fig. 1), and recorded extracellular, bilaterally synchronized respiratory rhythmic population activity. Inspiratory population activity was identified within the pre-Bötzinger complex⁹ (preBötC; Fig. 1a), a network that is necessary and sufficient for generating inspiration^{3,10}. Horizontal slices also generated postinspiratory population activity that discharged immediately after, but never during, inspiratory activity, and followed sighs generated in the preBötC¹¹ (Fig. 1a).

Postinspiratory bursts occurred spontaneously on average after 1 out of 12 preBötC bursts in the horizontal slice (Fig. 1a and Extended Data Fig. 2). The decreased excitability of postinspiratory activity compared to preBötC activity could be due to the absence of the pons, which, *in vivo*, provides descending neuromodulatory inputs, including norepinephrine, to the medulla¹². Indeed, postinspiratory activity in our slice preparation was exquisitely sensitive to norepinephrine. In the presence of 2 μ M norepinephrine, postinspiratory population activity

occurred with nearly every inspiratory cycle (Fig. 1a and Extended Data Fig. 2). Therefore, we used this norepinephrine concentration as a tool to facilitate postinspiratory activity *in vitro*.

Postinspiratory population activity was most pronounced approximately 400 μ m rostral to the preBötC, dorsal to the BötC, and caudal to the facial (VII) nucleus. We refer to this area as the postinspiratory complex (PiCo; Fig. 1a and Extended Data Fig. 1). To assess the distribution of postinspiratory activity, we positioned one electrode in the PiCo region and a second electrode contralaterally to map the amplitude of postinspiratory population activity across the VRC (Fig. 1a). Postinspiratory activity was concentrated rostral to the preBötC, but extended caudally and partially overlapped with inspiratory activity.

We anatomically characterized the PiCo by immunohistological labelling of transverse sections (Fig. 1b, left), revealing that choline acetyltransferase (ChAT)-positive cholinergic and vesicular glutamate transporter 2 (Vglut2; also known as Slc17a6)-expressing glutamatergic neurons in *Vglut2*-cre;Ai6 cre reporter mice co-localized in the PiCo, dorsomedial to the nucleus ambiguus (Fig. 1b, right). By contrast, ChAT-positive neurons in the nucleus ambiguus lacked substantial Vglut2 expression (Extended Data Fig. 3). In the sagittal plane, PiCo neurons that were co-labelled with ChAT and Vglut2 were located dorsal and caudal to the VII nucleus (Fig. 1c, left). Quantification of ChAT and Vglut2 co-expression revealed that the PiCo mainly extends from 40 to 280 μ m medial to the nucleus ambiguus and –50 to 250 μ m caudal to the VII nucleus caudal border (Fig. 1c, right). *In situ* hybridization confirmed that *Vglut2* mRNA was expressed in *Chat*-derived PiCo neurons in transverse sections from *Chat*-cre;Ai14 mice (Fig. 1d, e).

The Cre-dependent reporter line Ai27, which conditionally expresses channelrhodopsin-2 (ChR2) fused to td-Tomato in the presence of a selective, promoter-driven Cre, allowed photo-stimulation of specific neuronal sub-populations¹³. We recorded from PiCo neurons in *Vglut2*-cre;Ai27 and *Chat*-cre;Ai27 horizontal slices. Membrane depolarization of tetrodotoxin (TTX)-isolated PiCo neurons during light stimulation demonstrated that functionally identified postinspiratory cells were glutamatergic and cholinergic (Fig. 2a), consistent with the histological results. PiCo neurons generated neither pre-inspiratory bursts nor the biphasic discharge typical of pre-inspiratory neurons in the retrotrapezoidal nucleus parafacial respiratory group (RTN/pFRG) region¹⁴.

Postinspiratory population activity was unaffected by the use of bath-applied strychnine to block glycinergic inhibition (Fig. 2b and Extended Data Fig. 4). However, PiCo and preBötC bursts progressively synchronized following blockade of GABAergic inhibition with gabazine, in the presence (Fig. 2b and Extended Data Fig. 5) or absence (data not shown) of strychnine. The burst area of postinspiratory activity was increased when synaptic inhibition was blocked (Extended Data Fig. 4), indicating that the PiCo rhythm is modulated, but not generated, by inhibitory mechanisms. Inspiratory and postinspiratory bursts

¹Seattle Children's Research Institute, Center for Integrative Brain Research, Seattle, Washington 98101, USA. ²University of Washington School of Medicine, Graduate Program in Neuroscience, Seattle, Washington 98195, USA. ³University of Washington School of Medicine, Department of Neurological Surgery, Seattle, Washington 98105, USA.

*These authors contributed equally to this work.

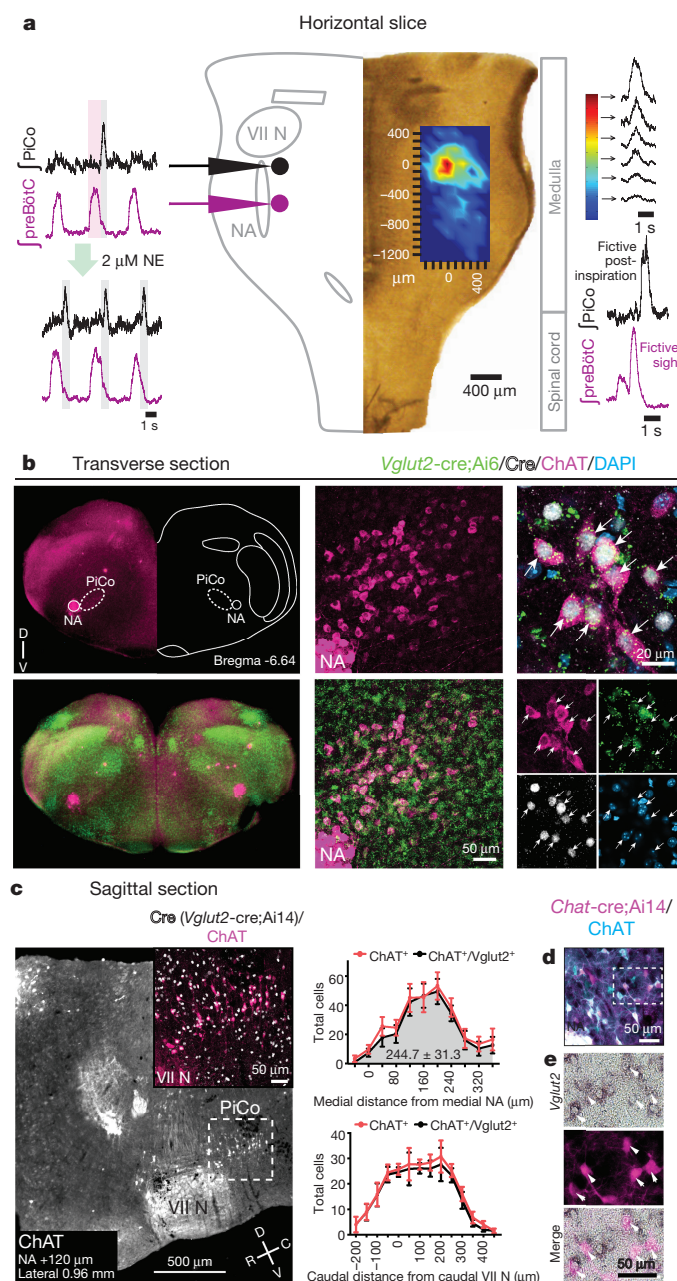


Figure 1 | Horizontal slice and anatomy of the PiCo. **a**, Population bursts in PiCo (black), preBötC (purple). Norepinephrine (NE, 2 μ M) stimulates PiCo bursts ($n = 23$). Left, schematic; right, heat map of PiCo burst amplitude ($n = 6$). PiCo bursts follow fictive sighs. **b**, Left, immunohistochemical labelling of PiCo dorsomedial to nucleus ambiguus (NA). Middle, higher magnification showing localization of ChAT and Vglut2 (ZsGreen1) within the PiCo of *Vglut2-cre;Ai6* mice. Right, arrows: higher magnification of triple-labelled ChAT⁺ ZsGreen1⁺ Cre⁺ PiCo cells (merged image above, individual images below; $n = 5$). **c**, Left, ChAT⁺ PiCo neurons (dashed box) in sagittal view. Inset, magnification of dashed box showing ChAT⁺ Cre⁺ PiCo neurons in *Vglut2-cre;Ai14* mice. Right, quantification of ChAT⁺ and Vglut2⁺ expressing PiCo cells mediolaterally from NA (top; 244.7 ± 31.3 average total cells; $n = 5$) and rostrocaudally from VII N (bottom; $n = 4$). Data shown as mean \pm s.e.m. **d**, ChAT⁺ PiCo neurons in *Chat-cre;Ai14* mice. **e**, Magnification of box in **d**. Arrowheads, *Vglut2* mRNA in *Chat*-derived PiCo neurons.

persisted following NMDA receptor blockade with 3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (CPP; data not shown), but bursting was abolished following non-NMDA glutamatergic blockade with 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; Fig. 2b). We conclude

that inspiratory and postinspiratory activities are generated by glutamatergic, non-NMDA dependent mechanisms, whereas the timing of inspiratory and postinspiratory bursts is established by GABAergic mechanisms.

Inspiratory rhythm-generating neurons in the preBötC are derived from *Dbx1*-expressing progenitor cells¹⁵. To investigate whether these neurons interact with PiCo, we used a tamoxifen-inducible transgenic line (*Dbx1-cre-ERT2;Ai27*) in which *Dbx1*-positive cells born after embryonic day (E)10.5 express channelrhodopsin-2. Photostimulation of preBötC neurons in horizontal slices from *Dbx1-cre-ERT2;Ai27* mice inhibited all recorded PiCo neurons in the presence of 2 μ M norepinephrine alone and after the addition of strychnine (Fig. 2c). In the presence of gabazine, this light-evoked inhibition was eliminated. The blockade of GABAergic inhibition revealed that PiCo neurons also received excitatory input from the preBötC, which was blocked by CNQX (Fig. 2c). Thus, inspiratory activity involving *Dbx1*-derived neurons simultaneously excites and inhibits PiCo neurons via glutamatergic and GABAergic mechanisms, respectively; however, under normal conditions, GABAergic input dominates over the concurrent glutamatergic excitation from the preBötC (Fig. 2c).

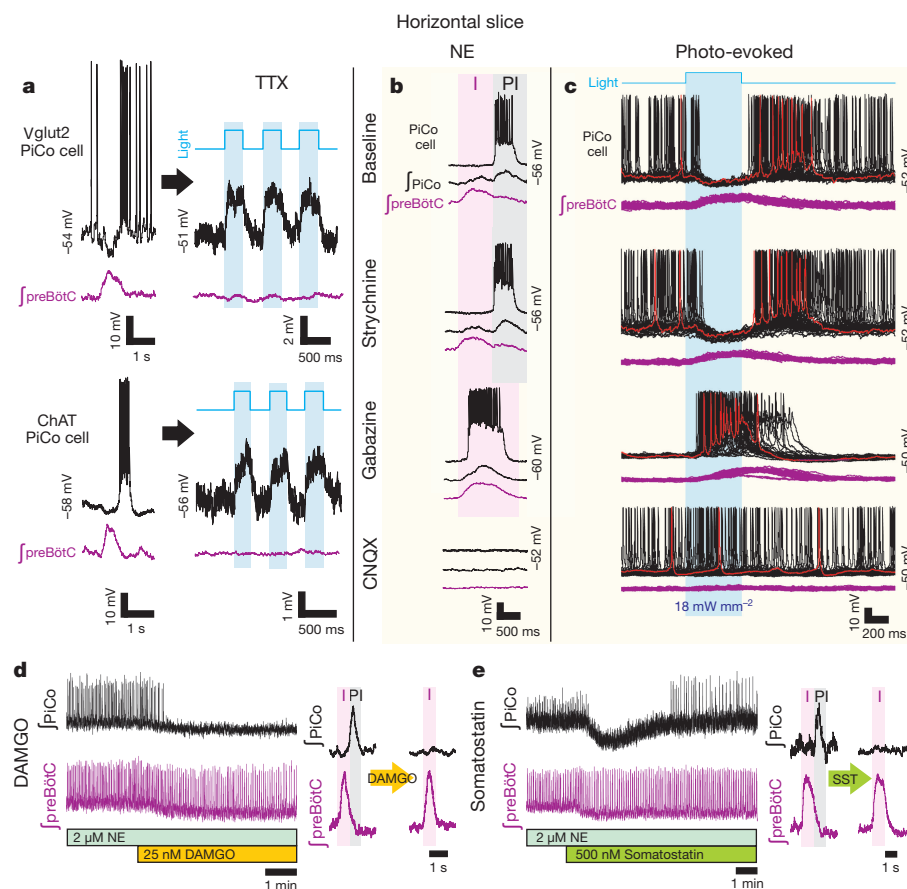
Because PiCo neurons co-express acetylcholine and glutamate (Fig. 1), we tested whether the postinspiratory rhythm depended on cholinergic mechanisms. Atropine, a muscarinic receptor antagonist, but not mecamylamine, a nicotinic receptor antagonist, depressed postinspiratory burst frequency. However, postinspiratory bursting persisted in the presence of both blockers and returned to near baseline frequency when the concentration of norepinephrine was increased to 4 μ M (Extended Data Fig. 6). Thus, the PiCo rhythm is modulated by, but not dependent on, cholinergic mechanisms.

PiCo neurons were intrinsically sensitive to the μ -opioid receptor agonist (D-Ala², N-Me-Phe⁴, Gly⁵-ol)-Enkephalin (DAMGO) the (Extended Data Fig. 7). In horizontal slices, PiCo population bursts were nearly eliminated by treatment with 25 nM DAMGO, whereas burst frequency in the preBötC was only slightly decreased (Fig. 2d and Extended Data Fig. 8). This exquisite opioid sensitivity unambiguously differentiates PiCo from the previously described RTN/pFRG, a region that is thought to contain the network that generates active expiration⁴ and is known to be insensitive to μ -opioid receptor activation¹⁶. The peptide somatostatin (SST) also inhibited postinspiratory PiCo activity, but had little effect on preBötC activity (Fig. 2e and Extended Data Fig. 8). These data are consistent with the inhibition of postinspiration by SST *in vivo*¹⁷.

Optogenetic stimulation of ChAT-positive neurons always elicited postinspiratory bursts from the PiCo in horizontal slices. Such stimulation never evoked an inspiratory burst (Extended Data Fig. 9) or burst activity in intracellularly recorded nucleus ambiguus neurons (data not shown). Because of this specificity for postinspiratory activity, we used adult *Chat-cre;Ai27* mice to stimulate the PiCo *in vivo*. Similar to the *in vitro* results, optogenetic activation of ChAT-positive neurons at the level of the PiCo (Fig. 3a–c) reliably evoked bursts in the cervical vagal nerve (cVN) (Fig. 3d and Extended Data Fig. 9). Photo-evoked postinspiratory bursts delayed the subsequent inspiration (Fig. 3d, e and Extended Data Fig. 9). This delay was eliminated by bilateral injection of DAMGO into the PiCo (Extended Data Fig. 10). Thus, postinspiration has a mutually inhibitory relationship with inspiratory activity.

To assess whether the PiCo is responsible for generating postinspiratory motor output *in vivo*, we took advantage of the sensitivity of the PiCo to SST and DAMGO. Bilateral injections of SST or DAMGO into the PiCo *in vivo* (Fig. 3a–c) markedly reduced the duration and amplitude of spontaneous vagal postinspiratory bursts (Fig. 3f–h). Collectively, these results suggest that the PiCo is both necessary and sufficient for generating postinspiratory activity.

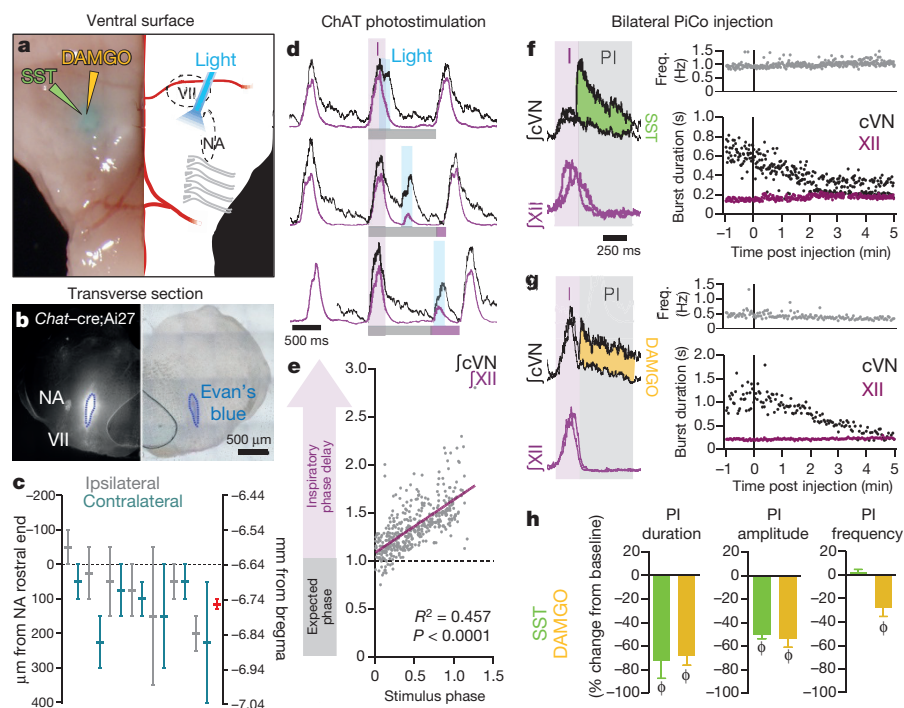
Moreover, the PiCo seems to possess autonomous rhythmogenic properties. In horizontal slices, norepinephrine concentrations above 2 μ M generated ectopic PiCo population bursts that outpaced the preBötC rhythm (Fig. 4a and Extended Data Fig. 2). Ectopic bursts



occurred in any phase of the inspiratory cycle except during preBötC bursts (Fig. 4b), consistent with photo-evoked PiCo bursts (Extended Data Fig. 9).

To further test the possibility that PiCo is an autonomous rhythm generator, we separated the VRC into two adjacent rostral and caudal transverse slices (Fig. 4c). Together, these slices span a total length of 1–1.1 mm of the rostrocaudal VRC, beginning with the caudal portion

of the VII nucleus. Recordings from the caudal face of each transverse slice revealed a slower rhythm in the rostral transverse slice containing the PiCo than in the caudal slice containing the preBötC^{3,11} (Fig. 4c). In the presence of 2 μ M norepinephrine, both transverse slices exhibited regular rhythmic activities with similar burst frequencies (Fig. 4c and Extended Data Fig. 2) that persisted with the progressive addition of strychnine, gabazine, and CPP, but were abolished by application of



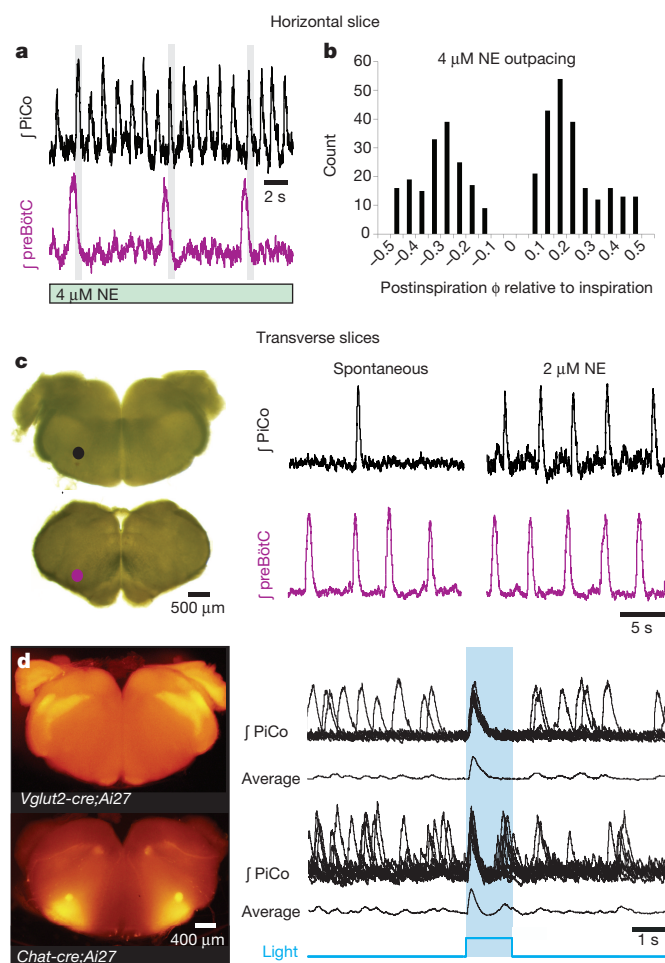


Figure 4 | The PiCo is an autonomous, rhythm-generating network. **a**, In the presence of 3–4 μM norepinephrine, the PiCo rhythm outpaces the preBötC rhythm, but the PiCo still produces bursts in the postinspiratory phase (grey bars) in horizontal slices. **b**, PiCo bursts occur in any phase except during inspiration (inspiratory peak = 0; count = 400 bursts; $n = 4$). **c**, Recordings from the PiCo and preBötC isolated in transverse slices *in vitro* slices; PiCo bursting is stimulated by 2 μM norepinephrine ($n = 33$). **d**, Light stimulation evokes a burst in rostral transverse slices from *Vglut2-cre;Ai27* mice ($91.3 \pm 5.1\%$ of stimulations, mean \pm s.e.m., $n = 4$) and *Chat-cre;Ai27* mice ($91.9 \pm 1.8\%$ of stimulations, mean \pm s.e.m., $n = 6$). 1.5-s light pulse. Top traces: 10 sweep overlay; bottom traces: sweep average.

CNXX (Extended Data Fig. 4), similar to the findings in horizontal slices. Consistent with our histological characterizations, optogenetic stimulation of either *Vglut2-cre;Ai27* or *Chat-cre;Ai27* rostral slices evoked population bursts in the PiCo (Fig. 4d). This provides additional evidence that glutamatergic–cholinergic neurons are important for rhythm generation within the PiCo network. Furthermore, PiCo activity was exquisitely sensitive to DAMGO and SST in isolated transverse slices (Extended Data Fig. 8). We conclude that the PiCo and preBötC can function as independent oscillators with similar rhythm-generating but distinct modulatory properties.

As an excitatory rhythmogenic network, the PiCo may not only be involved in the control of breathing, but might also contribute to the generation of other postinspiratory behaviours such as swallowing and vocalization. Although we did not perform behavioural assays in this study, various types of postinspiratory burst waveforms were observed in the vagal nerve (Extended Data Fig. 10) that were similarly affected by manipulation of the PiCo, supporting the idea that this network might have a broad role in controlling postinspiratory activities. In this context it will be interesting to resolve the role of the PiCo in specific

postinspiratory behaviours and to identify how the PiCo interacts with other neural networks such as the Kolliker-Fuse nucleus, a pontine structure that has been hypothesized to gate postinspiratory activity¹⁸, and the periaqueductal grey, a structure involved in vocalization and the control of postinspiration¹⁹.

On the basis of our results, we propose a triple oscillator model in which the three phases of breathing—inspiration, postinspiration, and active expiration—are generated by three spatially distinct excitatory rhythmogenic microcircuits, the preBötC, PiCo, and pFL, respectively, which are temporally coordinated by inhibitory interactions. The existence of discrete excitatory networks may facilitate the differential and dynamic control of ventilatory and non-ventilatory behaviours. Coupled oscillators²⁰ have also been hypothesized to be involved in networks controlling locomotion²¹, scratching²², swimming²³ and the circadian clock²⁴. Thus, this network organization may constitute a general principle of rhythm generation that promotes flexible control of complex biological processes.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 November 2015; accepted 17 June 2016.

Published online 27 July 2016.

- Richter, D. W. & Smith, J. C. Respiratory rhythm generation *in vivo*. *Physiology (Bethesda)* **29**, 58–71 (2014).
- Troche, M. S., Brandimore, A. E., Godoy, J. & Hegland, K. W. A framework for understanding shared substrates of airway protection. *J. Appl. Oral Sci.* **22**, 251–260 (2014).
- Smith, J. C., Ellenberger, H. H., Ballanyi, K., Richter, D. W. & Feldman, J. L. Pre-Bötzinger complex: a brainstem region that may generate respiratory rhythm in mammals. *Science* **254**, 726–729 (1991).
- Huckstepp, R. T., Cardoza, K. P., Henderson, L. E. & Feldman, J. L. Role of parafacial nuclei in control of breathing in adult rats. *J. Neurosci.* **35**, 1052–1067 (2015).
- Pagliardini, S. *et al.* Active expiration induced by excitation of ventral medulla in adult anesthetized rats. *J. Neurosci.* **31**, 2895–2905 (2011).
- Smith, J. C., Abdala, A. P., Rybak, I. A. & Paton, J. F. Structural and functional architecture of respiratory networks in the mammalian brainstem. *Phil. Trans. R. Soc. Lond. B* **364**, 2577–2587 (2009).
- Smith, J. C., Abdala, A. P., Koizumi, H., Rybak, I. A. & Paton, J. F. Spatial and functional architecture of the mammalian brain stem respiratory network: a hierarchy of three oscillatory mechanisms. *J. Neurophysiol.* **98**, 3370–3387 (2007).
- Alheid, G. F. & McCrimmon, D. R. The chemical neuroanatomy of breathing. *Respir. Physiol. Neurobiol.* **164**, 3–11 (2008).
- Ruangkittisakul, A., Kottick, A., Picardo, M. C., Ballanyi, K. & Del Negro, C. A. Identification of the pre-Bötzinger complex inspiratory center in calibrated “sandwich” slices from newborn mice with fluorescent Dbx1 interneurons. *Physiol. Rep.* **2**, e12111 (2014).
- Tan, W. *et al.* Silencing preBötzinger complex somatostatin-expressing neurons induces persistent apnea in awake rats. *Nat. Neurosci.* **11**, 538–540 (2008).
- Lieske, S. P., Thoby-Brisson, M., Telgkamp, P. & Ramirez, J. M. Reconfiguration of the neural network controlling multiple breathing patterns: eupnea, sighs and gasps. *Nat. Neurosci.* **3**, 600–607 (2000).
- Hilaire, G., Viemari, J. C., Coulon, P., Simonneau, M. & Bévengut, M. Modulation of the respiratory rhythm generator by the pontine noradrenergic A5 and A6 groups in rodents. *Respir. Physiol. Neurobiol.* **143**, 187–197 (2004).
- Madisen, L. *et al.* A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat. Neurosci.* **15**, 793–802 (2012).
- Animaru, H., Arata, A. & Homma, I. Firing properties of respiratory rhythm generating neurons in the absence of synaptic transmission in rat medulla *in vitro*. *Exp. Brain Res.* **76**, 530–536 (1989).
- Gray, P. A. *et al.* Developmental origin of preBötzinger complex respiratory neurons. *J. Neurosci.* **30**, 14883–14895 (2010).
- Janczewski, W. A. & Feldman, J. L. Distinct rhythm generators for inspiration and expiration in the juvenile rat. *J. Physiol. (Lond.)* **570**, 407–420 (2006).
- Burke, P. G., Abbott, S. B., McMullan, S., Goodchild, A. K. & Pilowsky, P. M. Somatostatin selectively ablates post-inspiratory activity after injection into the Bötzing complex. *Neuroscience* **167**, 528–539 (2010).
- Dutschmann, M. & Herbert, H. The Kolliker-Fuse nucleus gates the postinspiratory phase of the respiratory cycle to control inspiratory off-switch and upper airway resistance in rat. *Eur. J. Neurosci.* **24**, 1071–1084 (2006).
- Subramanian, H. H. & Holstege, G. Midbrain and medullary control of postinspiratory activity of the crural and costal diaphragm *in vivo*. *J. Neurophysiol.* **105**, 2852–2862 (2011).

20. Grillner, S. The motor infrastructure: from ion channels to neuronal networks. *Nat. Rev. Neurosci.* **4**, 573–586 (2003).
21. Kiehn, O. Locomotor circuits in the mammalian spinal cord. *Annu. Rev. Neurosci.* **29**, 279–306 (2006).
22. Stein, P. S. Neuronal control of turtle hindlimb motor rhythms. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **191**, 213–229 (2005).
23. Wiggin, T. D., Anderson, T. M., Eian, J., Peck, J. H. & Masino, M. A. Episodic swimming in the larval zebrafish is generated by a spatially distributed spinal network with modular functional organization. *J. Neurophysiol.* **108**, 925–934 (2012).
24. Evans, J. A., Elliott, J. A. & Gorman, M. R. Dynamic interactions between coupled oscillators within the hamster circadian pacemaker. *Behav. Neurosci.* **124**, 87–96 (2010).

Acknowledgements Supported by grants from the National Institute of Health NS087828-01 (to T.M.A.), HL090554 (to J.-M.R.), and HL126523-01 (to J.-M.R.). We thank P. Gray for insights provided throughout the preparation of this manuscript, T. Dashevskiy for creating the heat map of postinspiratory

activity, F. Bedogni for obtaining *in situ* hybridization reagents, and K. Cuthill for cryosectioning. J.-M.R. also thanks D. W. Richter and S. W. Schwarzacher for the inspiration to study postinspiration.

Author Contributions T.M.A., A.J.G., and J.-M.R. designed all experiments. T.M.A., A.J.G., N.A.B., J.P., J.C.B., A.D.W., and K.G.R. performed the experiments. T.M.A., A.J.G., N.A.B., and J.P. analysed the data. T.M.A., A.J.G., N.A.B., J.P., and J.-M.R. contributed to manuscript preparation. T.M.A. and J.-M.R. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-M.R. (nino1@uw.edu).

Reviewer Information *Nature* thanks J. Bouvier, O. Kiehn and H. Zoghbi for their contribution to the peer review of this work.

METHODS

Animals. All experiments were performed with the approval of the Institute of Animal Care and Use Committee of the Seattle Children's Research Institute. Mice were maintained with rodent diet and water available *ad libitum* in a vivarium with a 12-h light–dark cycle at 22 °C. In this study, we used both CD1 Swiss mice and Cre reporter mice generated on a C57BL/6 background. Ai27 mice were bred to conditionally express Channelrhodopsin-2 (H134R) fused to tdTomato inserted in the ROSA26 locus (B6.Cg–*Gt(ROSA)26Sor^{tm27.1(CAG-COP4::H134R/tdTomato)Hze}/J*; The Jackson Laboratory)²⁵. Ai6 mice were bred to express the green fluorescent protein ZsGreen1 inserted in the ROSA26 locus (B6.*Gt(ROSA)26Sor^{tm6(CAG-ZsGreen1)Hze}*; The Jackson Laboratory). Similarly, Ai14 mice were bred to express the red fluorescent protein tdTomato inserted into the ROSA26 locus (*Gt(ROSA)26Sor^{tm14(CAG-tdTomato)Hze}*; The Jackson Laboratory). Cre-driver mice expressed Cre recombinase under the control of subtype-specific promoters, including *Chat*–cre (B6;129S6–*Chat^{tm2(cre)Low}/J*; The Jackson Laboratory) and *Vglut2*–cre (B6;Slc17a6^{tm2(cre)Low}; Bradford Lowell). *Dbx1*–cre-ERT2 (*Dbx1^{CreERT2}* (refs 15, 26)) dams were bred with Ai27 males and pregnancies were timed and monitored. We intraperitoneally injected tamoxifen (25 mg/kg; from 10 mg tamoxifen (Sigma-Aldrich) dissolved per ml of corn oil) on embryonic day (E)10.5. Mice were typically born after 20 days of gestation. No method of randomization was used to determine how animals were allocated to experimental groups and the investigators were not blinded when analysing data in this study. No statistical methods were used to predetermine sample size.

In vitro slice preparations. We dissected the ventral respiratory column (VRC) using three types of brainstem slice: (1) A 'caudal' transverse slice that contains the preBötC as previously described^{3,11}; (2) a rostral transverse slice that encompasses the BötC and caudal portions of the VII nucleus; and (3) a horizontal slice that bilaterally isolates the VRC extending from the VII nucleus to the spinal cord. Slices were obtained at postnatal day (P)5–10 from CD1 and transgenic C57BL/6 mice. Animals were anaesthetized via rapid hypothermia on ice before quick decapitation at spinal cervical level C4–C5. The three slice types were differentiated by the cutting angle, plane and thickness of the slice.

For transverse slices, the head was pinned in a tissue-culture dish filled with a silicone elastomer (Sylgard). Skin and connective tissue were removed, and fine scissors were used to cut along skull sutures to separate the interparietal region of the skull and expose the cerebellum. A one-sided razor was used to make a single cut between the inferior colliculus and cerebellum. The brainstem was isolated by removing the cerebellum in ice-cold, oxygenated (95% O₂, 5% CO₂) artificial cerebrospinal fluid (aCSF) containing (in mM): 128 NaCl, 3 KCl, 1.5 CaCl₂, 1 MgCl₂, 24 NaHCO₃, 0.5 NaH₂PO₄, and 30 D-glucose (pH 7.4, 305–312 mOsm). A slanted (~15° from vertical) agar block was secured on a specimen tray, and then the isolated brainstem and spinal cord preparation was glued with cyanoacrylate to the slanted portion of the agar such that the rostral end was facing upward and the dorsal side was glued to the agar. Serial transverse slices proceeded on a vibratome until visual landmarks became clear. Once the 4th ventricle was completely open, a 550-µm slice was taken to obtain the rostral transverse slice containing the PiCo. The caudal face of this slice was characterized by containing the rostral-most portion of the nucleus ambiguus and the caudal portion of the VII nucleus. The subsequent 550-µm slice isolated the caudal transverse slice. This caudal slice was identical to the well-established transverse slice known to contain the preBötC¹¹. From an individual animal we routinely obtained both the rostral and the caudal slice preparation and recorded from the caudal side of each of the transverse slices in the same recording chamber.

To obtain the third type of slice preparation, the horizontal slice, the brainstem was mounted as described for the transverse slices. Serial coronal slices were taken from the rostral end of the brainstem until the facial nerves became visible, approximately 800–1,000 µm. The agar block was then removed from the specimen tray and reoriented so that the ventral surface of the brainstem faced upward and the blade advanced towards the rostral portion of the brainstem. The preparation was angled so that the ventral-most portion of the medulla was approximately level with the ventral-most portion of the spinal cord. The blade was positioned level to the rostroventral edge of the brainstem, stepped 900 µm downward (in the dorsal direction), and a single horizontal slice was cut retaining the ventral portion of the brainstem and spinal cord. The horizontal slice preserves long-range bilateral network interactions throughout the rostral-caudal axis of the VRC.

In vitro electrophysiology. All slices were immediately transferred to the recording chamber, where they were superfused with aCSF at a rate of 10 ml/min, bubbled continuously in carbogen (95% O₂ and 5% CO₂) to oxygenate and adjust pH to 7.4, and allowed to equilibrate to experimental temperature (33 ± 2 °C, thermoneutral zone for mice). Population activity was obtained by raising the extracellular potassium concentration from 3 mM to 8 mM in two steps over 30 min. This is defined as 'spontaneous conditions'.

Population activity was routinely recorded with borosilicate glass microelectrodes (World Precision Instruments) pulled on a Flaming/Brown micropipette puller (model P97, Sutter Instrument Co., <1 MΩ tip resistance) placed on the slice surface. Signals were amplified, filtered, and integrated as previously described²⁷. Automated burst analysis software was used to determine population recordings of burst frequency and amplitude²⁸.

The mapping experiment (Fig. 1a) was performed by placing a reference extracellular electrode on the PiCo and a second, mapping extracellular electrode on the contralateral side of a horizontal slice. The mapping electrode was systematically moved in 100-µm stereotaxic steps rostral, caudal, medial, and lateral to the PiCo. Postinspiratory burst amplitudes from the mapping electrode were normalized to that from the reference electrode to create a heat map of activity.

In horizontal and rostral transverse slices from transgenic mice expressing Chr2 in a subset of neurons, the PiCo (contralateral to the recording electrode) was light-stimulated using optical fibre (DPSSL Driver, blue 473 nm wavelength, 200 µm diameter, <22 mW/mm² intensity) for 500 ms or 1.5 s. Collections of 10 or 40 sweeps were recorded in succession (shown overlaid).

Intracellular blind patch recordings were performed on PiCo neurons. Borosilicate glass patch electrodes (with filaments, World Precision Instruments) were pulled (P-97 Flaming/Brown micropipette puller, Sutter Instrument Co.) to a 6–12 MΩ resistance. Electrodes were filled with an intracellular patch solution containing (in mM): 140 K-gluconic acid, 1 CaCl₂, 10 EGTA, 2 MgCl₂, 4 Na₂ATP, 10 HEPES (pH 7.4). Whole cell patch-clamp recordings were obtained in current clamp configuration using a Multiclamp 700B amplifier (Molecular Devices) sampling at 20 kHz. Extracellularly recorded signal was sampled at 1.67 kHz, amplified 10,000 times, filtered (low pass, 1.5 kHz; high pass, 250 Hz), rectified, and integrated using an electronic filter. Both extracellular and intracellular recordings were obtained with Clampex 10.0 (Molecular Devices). Recordings were stored on a computer for post-hoc analysis.

Receptor antagonists and neuromodulators were bath perfused during *in vitro* extracellular and intracellular recordings. All stock solutions were stored at –20 °C in small-volume aliquots to avoid repetitive freezing and thawing. Strychnine (1 µM, glycine receptor antagonist, Sigma Aldrich) and SR 95531 hydrobromide (gabazine, GABA_A receptor antagonist, 10 µM, Tocris) were used to block inhibitory synaptic transmission. To further block all fast synaptic transmission, 3-((±)2-carboxypiperazin-4-yl)propyl-1-phosphate (CPP, NMDA receptor antagonist, 10 µM, Tocris) and 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, AMPA receptor antagonist, 20 µM, Alimony Labs, diluted in DMSO) were bath applied. To block action potentials, 1 µM tetrodotoxin (TTX, Sigma-Aldrich) was used. To block cholinergic receptors, 10 µM atropine, a muscarinic receptor antagonist (Sigma-Aldrich) and 1 µM mecamylamine hydrochloride, a nicotinic receptor antagonist (Sigma Aldrich) were bath applied. To stimulate the PiCo rhythm, DL-norepinephrine hydrochloride (norepinephrine, 1–4 µM, Sigma-Aldrich) was used; and to inhibit the PiCo rhythm, [D-Ala², N-Me-Phe⁴, Gly⁵-ol]-Enkephalin (DAMGO, 1–300 nM, Sigma-Aldrich) and somatostatin (SST, 500 nM, Tocris) were applied.

In vivo electrophysiology. Adult mice were prepared as described previously²⁹. *Chat*–cre;Ai27 mice (P140–250) were anaesthetized with urethane (1.5 g/kg) and placed in a supine position, and the head was stabilized with ear bars. The trachea was exposed via a cervical midline incision and cannulated with a U-shaped tracheal tube. For the remainder of the surgery and experimental protocol, mice were allowed to spontaneously breathe humidified O₂ (FiO₂ = 100%). The rostral ends of the trachea and oesophagus were removed, followed by removal of the muscle and bone covering the ventral brainstem so that the vertebral and basilar arteries were visible. The dura and arachnoid membranes were removed followed by continuous perfusion of the ventral medullary surface with 95% O₂/5% CO₂ equilibrated aCSF solution at 37 ± 0.5 °C. The hypoglossal nerve (XII) and cervical vagus nerve (cVN) were isolated and cut, and their activity was measured using a suction electrode containing aCSF. Signals were amplified, bandpass filtered (8 Hz–3 kHz), and digitized with a Digidata 1400 and pClamp 10 software (Molecular Devices).

After completion of the surgery, mice were allowed to stabilize for 15 min before 15–20 min of baseline respiratory activity was recorded. Using the vertebral and basilar arteries as landmarks (Fig. 3a–c), 200 µm diameter optical fibres coupled to a 447 nm DPSSL driver laser at intensity <230 mW/mm² were placed bilaterally on the ventral surface of the medulla above the region containing the PiCo. XII and cVN activity was recorded during 10-s episode files containing a 200-ms light pulse to stimulate *Chat*–cre-expressing cells. Inspiratory and postinspiratory (spontaneous or light-evoked) activity was analysed using Clampfit 10 software (Molecular Devices). The phase of evoked cVN postinspiratory activity was determined as the fraction of the inspiratory cycle (Extended Data Fig. 9) or the fraction of the average duration of the preceding two inspiratory cycles (expected phase, Fig. 3e). The inspiratory phase duration during cycles containing a light-evoked PiCo burst were then divided by the expected phase to determine the phase delay (Fig. 3e).

In some mice, the PiCo was photostimulated before and after bilateral injection of 5 μ M DAMGO to assess the effect of DAMGO on the inspiratory phase delay (Extended Data Fig. 10).

To test whether PiCo activity is necessary for postinspiration *in vivo*, pulled micropipettes containing somatostatin (SST; 750 μ M) or DAMGO (5 μ M), and either Evan's Blue or Fast Green to identify the injection site, were inserted (300–400 μ m) bilaterally into the PiCo. Spontaneous postinspiratory burst amplitude, duration, and frequency were then quantified 3–5 min after a 250-nL injection of SST or DAMGO and compared (using student's *t*-test) to pre-injection values (GraphPad, Prism 5 software). Postinspiratory burst duration was determined by subtracting the duration of XII nerve inspiratory activity from the duration of the corresponding cVN burst. Postinspiratory amplitude was defined as the amplitude of cVN activity immediately following XII nerve inspiratory activity. Following experimental protocols, mice were perfused with 4% paraformaldehyde (PFA), and brainstems were extracted and cryoprotected (30% sucrose in PBS). Brainstems were then serially sectioned to identify sites of injection.

Immunohistochemistry. 200- μ m rostral transverse slices from *Vglut2-cre*;Ai14 mice were fixed with 4% PFA for 1 h and immunostained as whole mounts. Slices were washed in PBST (0.1–0.5% Triton X-100), blocked with 10% donkey serum in PBST overnight at 4°C, incubated for 2–3 days in primary antibody in blocking solution at 4°C, washed in PBST, incubated in secondary antibody in blocking solution for 5–8 h at room temperature, washed in PBST, counterstained with 0.01% DAPI (Life Technologies), and mounted in Fluoromount-G (SouthernBiotech). 40- μ m sagittal sections from *Vglut2-cre*;Ai14 mice were also immunolabelled for quantification. Mice were transcardially perfused with 4% PFA, and brainstems were postfixed in 4% PFA overnight. Isolated brainstems were transferred through increasing sucrose gradients (10–30%), embedded in OCT compound (TissueTek), frozen, and cryosectioned. Immunohistochemical labelling followed the same protocol as for whole mounts with shortened incubation times. Primary antibodies included anti-ChAT (1:100, AB144P, Millipore) and anti-Cre recombinase (1:200, 908001, BioLegend). Secondary antibodies were Alexa Fluor 568- or 647-conjugated (1:250, Life Technologies). Maximum intensity projections of optical slice z-stacks were acquired using a Zeiss 710 Quasar 34-channel LSCM (Carl Zeiss). Cre-labelled images were despeckled for background noise reduction.

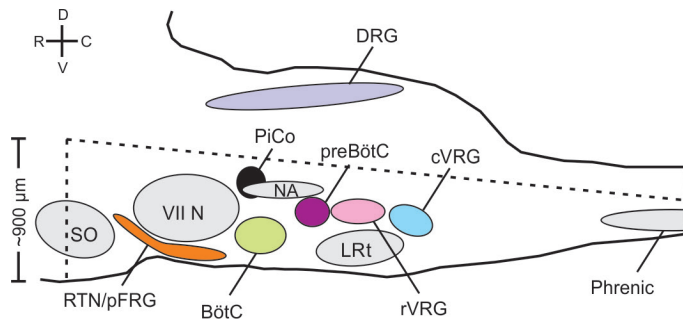
Cell counting. Maximum intensity projections of 20 \times optical slice z-stacks were collected –40 to 360 μ m medial to the medial end of the nucleus ambiguus. ChAT⁺ and Cre⁺ or tdTomato⁺ cells were counted within this area, with the exception of ChAT⁺ cells that clearly belonged to the nucleus ambiguus or VII nucleus (distinguished by location and large cell size). Counts from each hemisphere were averaged for individual animals. Counts in the rostrocaudal direction were taken from sagittal slices 40–280 μ m medial to the medial end of nucleus ambiguus, where PiCo cells are most abundant. ChAT⁺ and Cre⁺ or tdTomato⁺ cells were counted in 50- μ m bins through the rostrocaudal extent and summed across the 240- μ m span. For *in vivo* experiments, 50- μ m serial transverse brainstem sections were cryopreserved and processed as described above. Sections were imaged through the region encompassing the injection site and noted for the presence of Evan's blue or Fast Green dye. *Chat-cre* expression identified the rostral end of the nucleus ambiguus in order to quantify the rostrocaudal location of injection sites relative

to the nucleus ambiguus. Anatomical diagrams and coordinates were adapted from the Franklin and Paxinos adult mouse brain atlas³⁰.

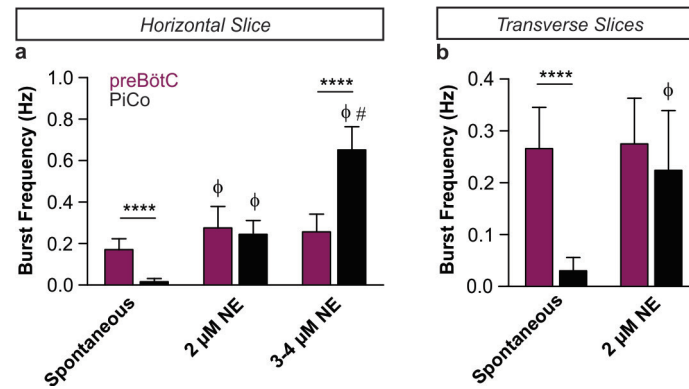
In situ hybridization. P8–11 mice were perfused with 4% PFA (0.1 M sodium phosphate, pH 7.0), and brainstems were post-fixed in 4% PFA (0.1 M sodium phosphate, pH 7.0) + 4% sucrose, overnight at 4°C. Isolated brainstems were submerged in 30% sucrose, embedded in OCT, frozen at –80°C, and cryosectioned at 20 μ m. Prior to hybridization, sections were fixed with 4% PFA/DEPC-PBS, pH 7.0 at 4°C for 5 min, treated with proteinase K (1 μ g/ml) for 10 min at room temperature, fixed with 4% PFA/DEPC-PBS, pH 7.0 at 4°C for 5 min, and acetylated for 10 min at room temperature. DIG-labelled *Vglut2*-Dig antisense RNA probe (306 bp fragment of *Vglut2* (1563–1869 bp, XM_006540602)) was hybridized onto sections (0.8 μ g/ml) at 42°C overnight. Following hybridization, sections were incubated with RNase A (50 μ g/ml, Invitrogen) for 30 min at 37°C. DIG Nucleic acid detection kit (Roche) was used for RNA probe detection. The sections were incubated in anti-digoxigenin-AP conjugate (Roche Applied Science, sheep, 1:1,000) for 1 h at room temperature. Hybridized molecules were visualized after incubation in an enzyme-catalysed colour reaction with a solution of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitroblue tetrazolium salt (NBT) (Roche Applied Science). The sections were developed in the BCIP/NBT solution for 2 h in the dark at room temperature. The enzyme-catalysed colour reaction was stopped with TE, pH 8.0 and fixed in 4% PFA, pH 7.0 for 20 min at 4°C.

Statistics. All statistics were performed using GraphPad Prism 5. Numerical data are reported as the mean \pm s.e.m. Normality was determined by D'Agostino-Pearson normality test. For normally distributed data, statistical significance was assessed by two-tailed paired Student's *t*-tests and two-way ANOVAs where appropriate. Two-way ANOVAs were followed by Bonferroni post-hoc correction. For data that were not normal, we used non-parametric two-tailed Mann-Whitney, Kruskal-Wallis, one-way ANOVA, and repeated measures Friedman tests where appropriate. Kruskal-Wallis and Friedman tests were followed by Dunn's multiple comparison post-hoc tests. Variance was similar between groups that were statistically compared. Results were considered significant when $P < 0.05$. α was set less than or equal to 0.05 for multiple comparison tests. Sample sizes were chosen on the basis of previous studies.

25. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).
26. Hirata, T. *et al.* Identification of distinct telencephalic progenitor pools for neuronal diversity in the amygdala. *Nat. Neurosci.* **12**, 141–149 (2009).
27. Hill, A. A., Garcia, A. J., Ill, Zanella, S., Upadhyaya, R. & Ramirez, J. M. Graded reductions in oxygenation evoke graded reconfiguration of the isolated respiratory network. *J. Neurophysiol.* **105**, 625–639 (2011).
28. Telgkamp, P., Cao, Y. Q., Basbaum, A. I. & Ramirez, J. M. Long-term deprivation of substance P in PPT-A mutant mice alters the anoxic response of the isolated respiratory network. *J. Neurophysiol.* **88**, 206–213 (2002).
29. Doi, A. & Ramirez, J. M. State-dependent interactions between excitatory neuromodulators in the neuronal control of breathing. *J. Neurosci.* **30**, 8251–8262, (2010).
30. Franklin, K. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates, Third Edition* (Academic, 2008).

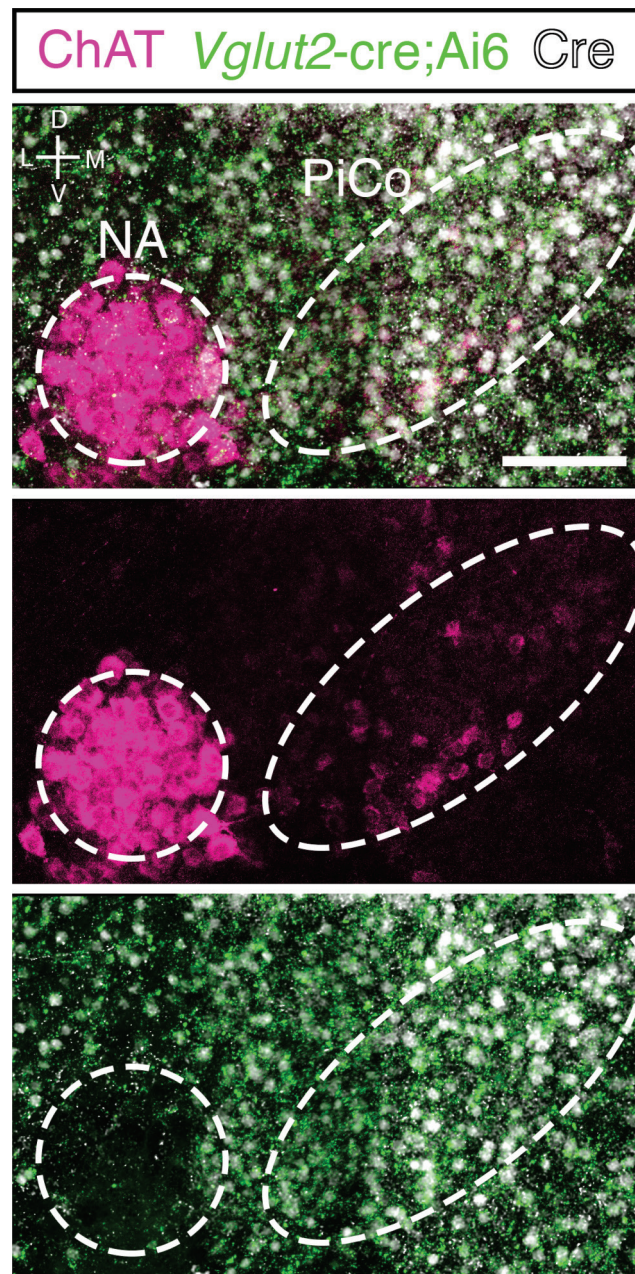


Extended Data Figure 1 | Schematic of the horizontal slice from a sagittal view that retains the medullary ventral respiratory column in the brainstem. Dotted lines represent approximate boundaries of the horizontal slice. Slice retains part of the superior olive (SO), and the entire retrotrapezoidal nucleus/para-facial respiratory group (RTN/pFRG), facial nucleus (VII N), Bötzinger complex (BötC), postinspiratory complex (PiCo), nucleus ambiguus (NA), preBötzinger complex (preBötC), lateral reticular nucleus (LRt), and the rostral and caudal ventral respiratory groups (rVRG and cVRG, respectively). The slice also retains a portion of the spinal cord and includes part of the phrenic motor nucleus (approximately cervical segments 3 and 4). The slice does not contain the dorsal portion of the medulla including the dorsal respiratory group (DRG). Dorsal (D), ventral (V), rostral (R), caudal (C).

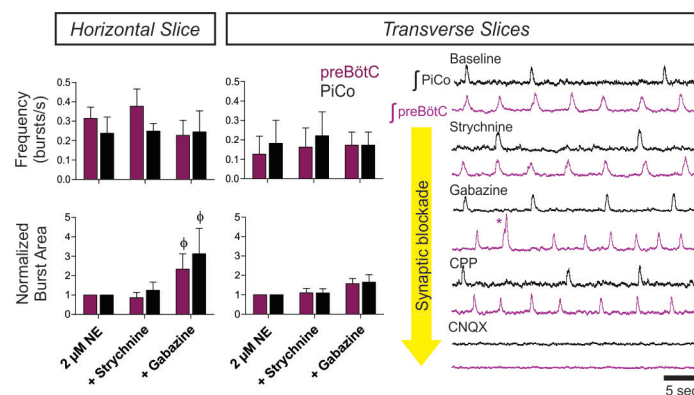


Extended Data Figure 2 | Norepinephrine dose response of preBötC and PiCo rhythms in horizontal and transverse slices. The frequency of the PiCo rhythm (black) is highly sensitive to the application of low concentrations of norepinephrine while the preBötC rhythm (purple) stays relatively constant in both types of slice preparations. **a**, In horizontal slices the PiCo rhythm is slow under spontaneous conditions ($n = 10$), the two rhythms have similar burst frequencies in the presence of 2 μ M norepinephrine ($n = 6$), and the PiCo rhythm significantly outpaces the

preBötC rhythm under higher concentrations of norepinephrine ($n = 4$, 3–4 μ M norepinephrine). **b**, Similarly, when isolated in transverse slices, the PiCo rhythm has a slow frequency under spontaneous conditions ($n = 10$), and the preBötC and PiCo have similar frequencies at 2 μ M norepinephrine ($n = 7$; mean \pm s.e.m.). Two-way ANOVA followed by a Bonferroni post-hoc test. **** $P < 0.0001$ comparing PiCo to preBötC, $\phi P < 0.05$ compared to baseline (Spon.), $\#P < 0.05$ compared to 2 μ M norepinephrine.

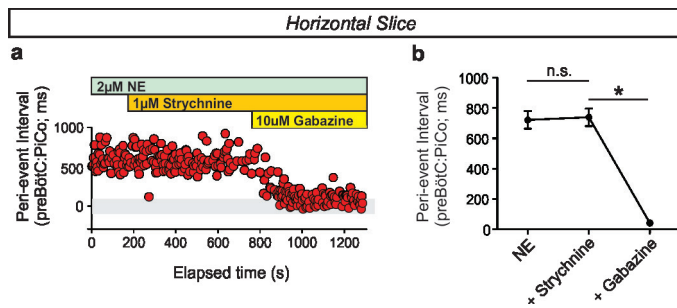


Extended Data Figure 3 | Nucleus ambiguus neurons lack *Vglut2-cre* expression. High magnification view at the level of PiCo from a *Vglut2-cre;Ai6* (*ZsGreen1*; green) mouse immunolabelled with ChAT antibody (magenta) and Cre antibody (white). Note lack of green *Vglut2-cre* expression in the nucleus ambiguus. Scale bar, 100 μ m.



Extended Data Figure 4 | Progressive synaptic blockade in horizontal and transverse slices. Left, graphs comparing frequency and normalized burst area between horizontal ($n = 5$) and paired transverse slices ($n = 5$) after the application of strychnine and gabazine. In both horizontal and paired transverse slices, PiCo and preBötC rhythms have nearly identical burst frequencies in the presence of gabazine (top). The burst area of both rhythms also significantly increases with the application of gabazine in both slice preparations (bottom). Two-way ANOVA followed

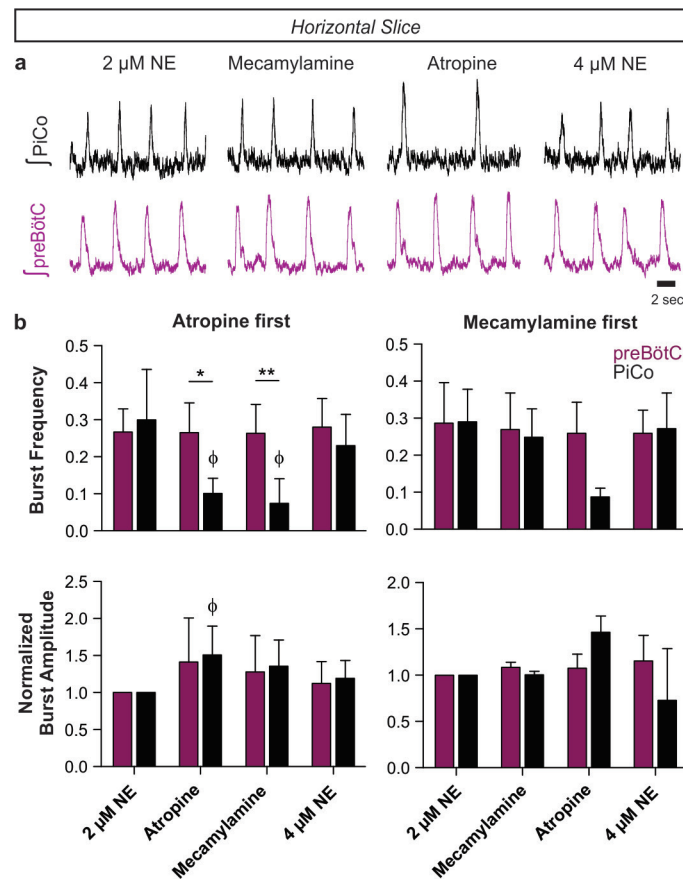
by a Bonferroni post-hoc test. $^{\phi}P < 0.05$ compared to baseline (2 μ M norepinephrine). Right, synaptic blockers were progressively perfused over paired transverse slices at 10-min intervals. Both PiCo and preBötC rhythms persist in the presence of 1 μ M strychnine, 10 μ M gabazine, and 10 μ M CPP. Population rhythms ceased in the presence of 20 μ M CNQX, indicating that both rhythms are excitatory ($n = 5$). The asterisk denotes a characteristic sigh in the preBötC trace.



Extended Data Figure 5 | Peri-event interval between preBötC and PiCo bursts during inhibitory block in the horizontal slice. **a**, Peri-event interval, time between peak of preBötC and PiCo bursts, is constant in strychnine; however, gabazine initiates progressive synchronization between rhythms (shown here in a representative experiment).

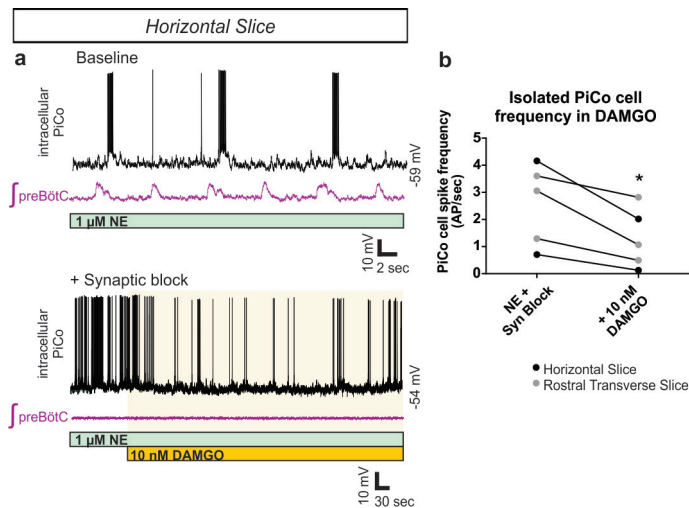
b, Average peri-event intervals at baseline and after sequential application of strychnine and gabazine ($n = 6$, mean \pm s.e.m.). Repeated measures Friedman test followed by Dunn's multiple comparisons post-hoc test.

* $P < 0.05$.

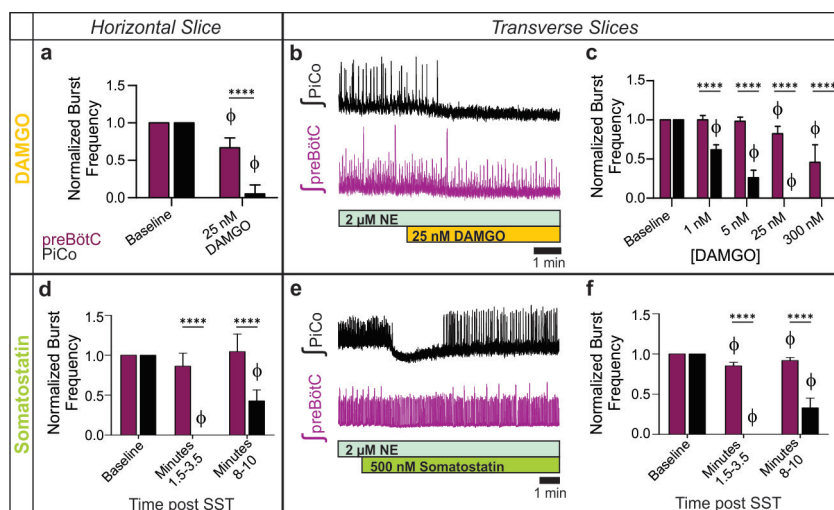


Extended Data Figure 6 | Blocking muscarinic and nicotinic acetylcholine receptors does not abolish the PiCo rhythm. **a**, Raw population bursts from PiCo and contralateral preBötC with the progressive addition of 1 μ M mecamylamine (nicotinic receptor antagonist), 10 μ M atropine (muscarinic receptor antagonist), and 4 μ M norepinephrine. **b**, The left two graphs show $n = 5$ experiments in which atropine was applied first, and the right graphs illustrate $n = 3$ experiments in which mecamylamine was applied first. Blockade of muscarinic receptors results in a larger decrease in PiCo burst frequency than

blocking nicotinic receptors, while preBötC frequency does not change significantly (top graphs). Blockade of muscarinic receptors increases the amplitude of PiCo bursts (bottom graphs). The PiCo rhythm persists after concurrent blockade of both types of acetylcholine receptors, and PiCo burst frequency rebounds to near baseline levels when an additional 2 μ M norepinephrine is applied (total 4 μ M norepinephrine; top graphs; mean \pm s.e.m.). Two-way ANOVA followed by a Bonferroni post-hoc test. ** $P < 0.01$, * $P < 0.05$ comparing preBötC to PiCo, $\phi P < 0.05$ compared to baseline (2 μ M norepinephrine).

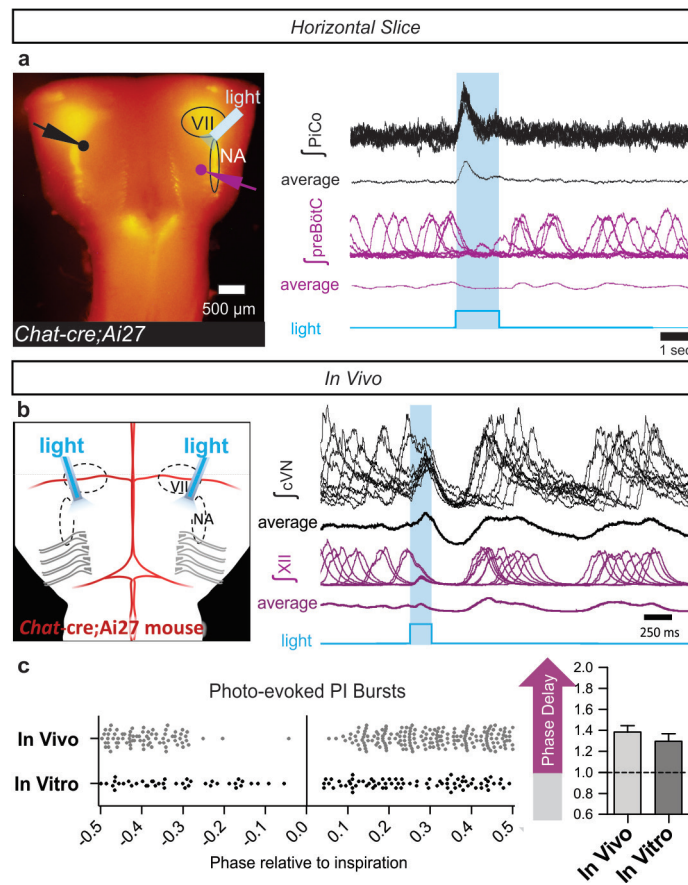


Extended Data Figure 7 | Synaptically isolated PiCo neurons decrease firing frequency in the presence of DAMGO. **a**, Top traces show intracellular recordings from PiCo cells with concurrent extracellular preBötC population activity from a horizontal slice under 1 μ M norepinephrine baseline conditions. Bottom traces show the same recordings after blocking fast synaptic transmission (1 μ M strychnine, 10 μ M gabazine, 10 μ M CPP, 20 μ M CNQX) to synaptically isolate PiCo neurons. Application of 10 nM DAMGO decreases the cell's intrinsic firing frequency. **b**, Quantified data show that DAMGO significantly decreases action potential (AP) firing frequency of synaptically isolated PiCo neurons in both horizontal slices (black dots) and transverse PiCo slices (grey dots) (two-tailed paired *t*-test, **P* < 0.05; *n* = 5).



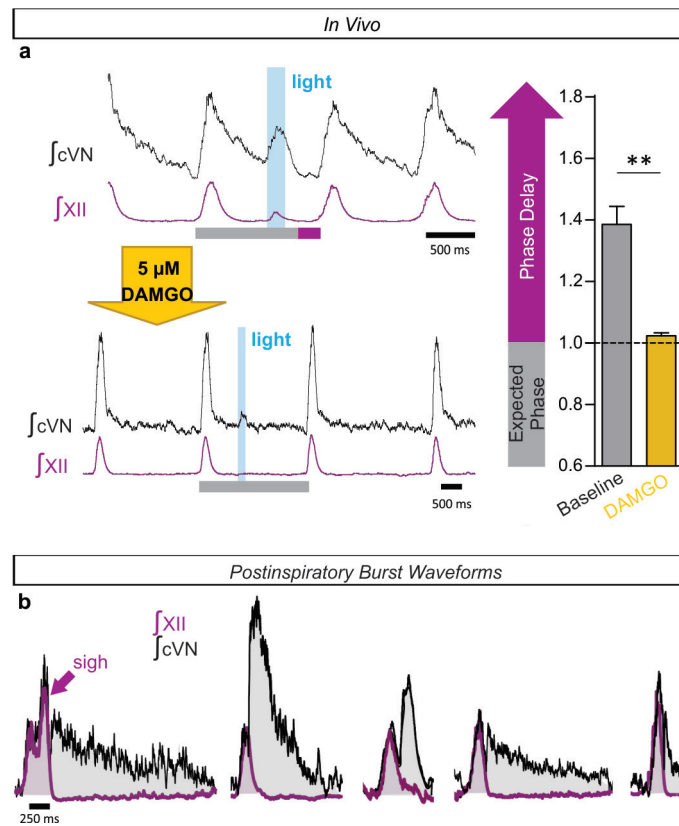
Extended Data Figure 8 | Differential PiCo and preBötC population responses to DAMGO and SST in horizontal and transverse slices. **a**, After the application of 25 nM DAMGO, preBötC burst frequency decreases only slightly ($n = 5$), whereas PiCo bursting is nearly eliminated. **b**, Similar to results observed in horizontal slices, the PiCo rhythm is eliminated by 25 nM DAMGO in transverse slices that isolate PiCo and preBötC in the presence of 2 μ M norepinephrine ($n = 5$). Periodic large amplitude bursts in the bottom preBötC trace are fictive sighs. **c**, DAMGO dose response of normalized preBötC and PiCo burst frequency in transverse slices, illustrating the differential sensitivity of the PiCo and preBötC rhythms to DAMGO; burst frequency values are normalized to baseline frequency in 2 μ M norepinephrine (mean \pm s.e.m., $n = 8$ with

minimum replicates of 4 for each location and concentration). **d**, The PiCo rhythm is selectively and transiently inhibited by the application of 500 nM SST whereas the preBötC rhythm persists in horizontal slices. Graph shows normalized average burst frequencies of both rhythms at baseline, 1.5–3.5 min after SST application, and 8–10 min after SST application ($n = 6$). **e**, Similar to the horizontal slice, SST application results in a robust inhibition of PiCo bursting in paired transverse slices. **f**, Similar to **d**, compiled normalized burst frequencies for $n = 6$ transverse slices before and after SST application. (mean \pm s.e.m.) Two-way ANOVA followed by a Bonferroni post-hoc test. **** $P < 0.0001$ comparing PiCo to preBötC, $\phi P < 0.05$ compared to baseline.



Extended Data Figure 9 | Light stimulation of cholinergic cells evokes postinspiratory activity in horizontal slices and *in vivo*. **a**, Two population electrodes were placed at the level of PiCo (black dot and trace) and contralateral preBötC (purple dot and trace) in a horizontal slice from a *Chat-cre;Ai27* mouse. Under spontaneous conditions (no norepinephrine), cholinergic neurons expressing channelrhodopsin-2 were light activated with an optical fibre (labelled 'light') placed over the PiCo ipsilateral to the preBötC electrode. PiCo population bursts were triggered upon the onset of a 1.5-s light pulse whereas no bursts were light

evoked in the preBötC ($n = 6$). Figure shows 10 traces overlaid for each electrode with averaged traces below from a representative experiment. **b**, Photo-stimulating PiCo in adult anaesthetized *Chat-cre;Ai27* mice reliably triggers cVN bursts. Figure shows 10 traces overlaid with averages below of cVN and XII activity during a 200-ms light stimulation of PiCo. **c**, Postinspiratory bursts can be photo-evoked both *in vivo* ($n = 6$) and *in vitro* ($n = 6$) at any phase except for during inspiration and just before inspiration (bottom left) owing to the inspiratory phase delay that occurs when PiCo is stimulated (mean \pm s.e.m., bottom right).



Extended Data Figure 10 | Elimination of phase delay by DAMGO and diversity of postinspiratory waveforms *in vivo*. **a**, Injection of 5 μ M DAMGO into the PiCo eliminates the phase delay elicited by photostimulation of PiCo in *Chat-cre;Ai27* mice. A representative experiment showing cVN and XII recordings during a 200-ms light pulse before and after injection of PiCo with DAMGO (left; grey bars, expected phase; purple bars, inspiratory phase delay) and the average inspiratory

phase delay (right) (mean \pm s.e.m., two-tailed paired *t*-test, $**P < 0.01$; $n = 6$). **b**, Diverse postinspiratory vagal waveforms were recorded *in vivo*. Five examples of cVN (black) and XII (purple) recordings (overlaid) show that postinspiratory activity can vary from large decrementing patterns to small short bursts, potentially representing the neural basis of a variety of postinspiratory behaviours.

Synchronized cycles of bacterial lysis for *in vivo* delivery

M. Omar Din^{1*}, Tal Danino^{2†*}, Arthur Prindle¹, Matt Skalak², Jangir Selimkhanov¹, Kaitlin Allen², Ellixis Julio¹, Eta Atolia², Lev S. Tsimring³, Sangeeta N. Bhatia^{2,4,5,6,7,8§} & Jeff Hasty^{1,3,9§}

The widespread view of bacteria as strictly pathogenic has given way to an appreciation of the prevalence of some beneficial microbes within the human body^{1–3}. It is perhaps inevitable that some bacteria would evolve to preferentially grow in environments that harbour disease and thus provide a natural platform for the development of engineered therapies^{4–6}. Such therapies could benefit from bacteria that are programmed to limit bacterial growth while continually producing and releasing cytotoxic agents *in situ*^{7–10}. Here we engineer a clinically relevant bacterium to lyse synchronously at a threshold population density and to release genetically encoded cargo. Following quorum lysis, a small number of surviving bacteria reseed the growing population, thus leading to pulsatile delivery cycles. We used microfluidic devices to characterize the engineered lysis strain and we demonstrate its potential as a drug delivery platform via co-culture with human cancer cells *in vitro*. As a proof of principle, we tracked the bacterial population dynamics in ectopic syngeneic colorectal tumours in mice via a luminescent reporter. The lysis strain exhibits pulsatile population dynamics *in vivo*, with mean bacterial luminescence that remained two orders of magnitude lower than an unmodified strain. Finally, guided by previous findings that certain bacteria can enhance the efficacy of standard therapies¹¹, we orally administered the lysis strain alone or in combination with a clinical chemotherapeutic to a syngeneic mouse transplantation model of hepatic colorectal metastases. We found that the combination of both circuit-engineered bacteria and chemotherapy leads to a notable reduction of tumour activity along with a marked survival benefit over either therapy alone. Our approach establishes a methodology for leveraging the tools of synthetic biology to exploit the natural propensity for certain bacteria to colonize disease sites.

In order to control population levels and facilitate drug delivery using bacteria, we engineered a synchronized lysis circuit (SLC) using coupled positive and negative feedback loops that have previously been used to generate robust oscillatory dynamics^{12,13}. The circuit (Fig. 1a) consists of a common promoter that drives expression of both its own activator (positive feedback) and a lysis gene (negative feedback). Specifically, the *luxI* promoter regulates production of autoinducer (AHL), which binds LuxR and enables it to transcriptionally activate the promoter. Negative feedback arises from cell death that is triggered by a bacteriophage lysis gene (ϕ X174 E) which is also under control of the *luxI* promoter^{13–15}. AHL can diffuse to neighbouring cells and thus provides an intercellular synchronization mechanism.

The bacterial population dynamics arising from the synchronized lysis circuit can be conceptualized as a slow build-up of the signalling

molecule (AHL) to a threshold level, followed by a lysis event that rapidly prunes the population and enables the release of bacterial contents (Fig. 1b). After lysis, a small number of remaining bacteria begin to produce AHL anew, allowing the ‘integrate and fire’ process to be repeated in a cyclical fashion. We used microfluidic devices to observe growth and lysis with the fluorescent protein superfolder GFP (sfGFP) as a proxy for circuit dynamics in attenuated *Salmonella enterica* subsp. *enterica* serovar Typhimurium (Supplementary Videos 1 and 2). We observed periodic lysis events characterized by peaks in the fluorescent reporter expression that correspond to population lysis (Fig. 1c). The fraction of lysed cells remains consistent across subsequent cycles, suggesting that lysis and survival occur in a stochastic manner (Extended

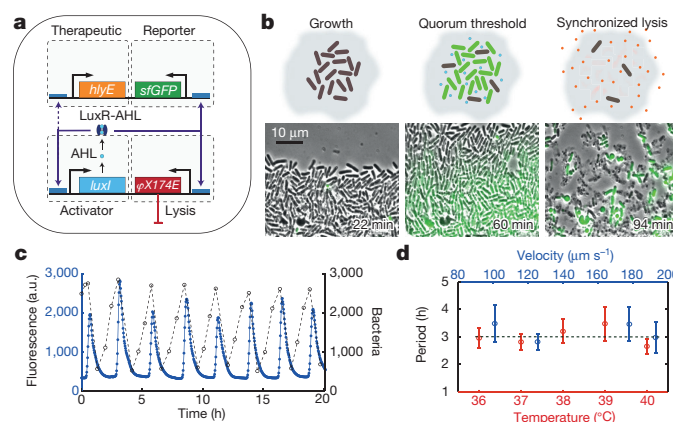


Figure 1 | Construction and characterization of the SLC. a, The circuit contains an activator¹³ and lysis plasmid. When the population reaches the quorum threshold at a critical AHL concentration, the *luxI* promoter drives the transcription of gene *E* for lysis, *luxI*, and *sfGFP* or *luxCDABE* as the reporter module. The *luxI* or the *tac* promoter also drives the transcription of the therapeutic gene for the stabilized circuit used *in vivo*. LuxR in this system is driven by the native *luxR* promoter. **b**, The main stages of each lysis cycle from seeding to quorum ‘firing’. Shown below the schematic depictions are typical time series images of the circuit-harboring cells undergoing the three main stages of quorum firing in a microfluidic growth chamber¹². **c**, Fluorescence profile of a typical microfluidic experiment. The estimated cell population trajectory reveals that lysis events correspond to peaks of sfGFP fluorescence. **d**, Period as a function of estimated flow velocity in the media channel of the microfluidic device and environmental temperature. Error bars indicate ± 1 s.d. for 13–50 peaks. These experiments were performed with strain 1, see Supplementary Information for complete strain information.

¹Department of Bioengineering, University of California, San Diego, La Jolla, California 92093, USA. ²Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³BioCircuits Institute, University of California, San Diego, La Jolla, California 92093, USA. ⁴Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02139, USA. ⁵Department of Medicine, Brigham and Women’s Hospital, Boston, Massachusetts 02139, USA. ⁶Electrical Engineering and Computer Science and David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁷Marble Center for Cancer Nanomedicine and Ludwig Center for Molecular Oncology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁸Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. ⁹Molecular Biology Section, Division of Biological Science, University of California, San Diego, La Jolla, California 92093, USA. [†]Present address: Department of Biomedical Engineering, Columbia University, New York, New York 10027, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

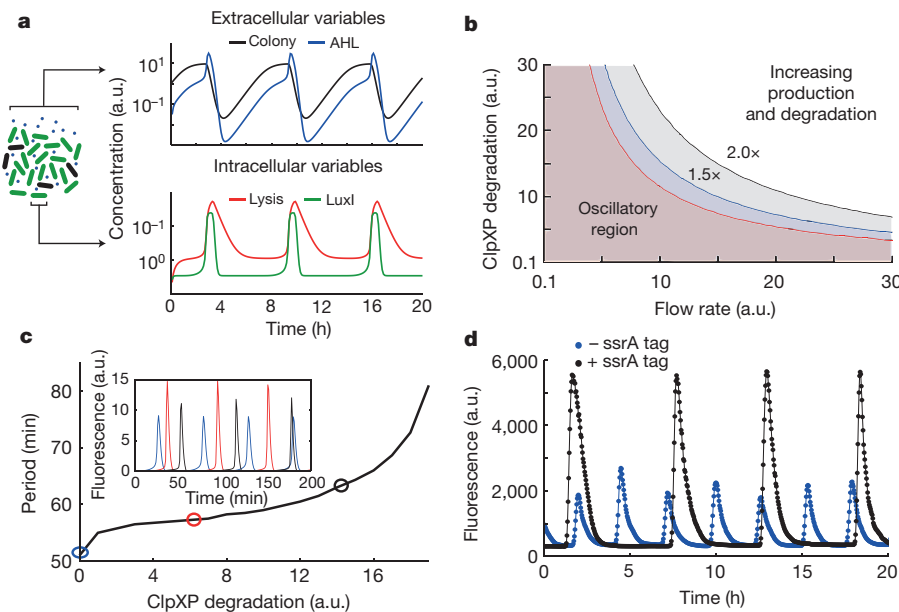


Figure 2 | Computational modelling and tunability. **a**, The model consists of intracellular variables (lysis protein E and LuxI concentrations) and extracellular variables (colony size and AHL concentrations). A time series of colony size (black), colony AHL (blue), intracellular LuxI (green) and lysis protein concentrations (red) are shown on the right. **b**, The region in the model parameter space for ClpXP-mediated degradation (see Supplementary Information) and flow where the model output is oscillatory increases with higher production and degradation terms. **c**, Results from the computational model showing the ability to tune the oscillatory period by varying ClpXP mediated degradation of LuxI. **d**, Fluorescence profiles showing lysis oscillations for LuxI ssrA (black, strain 2) and LuxI non-ssrA (blue, strain 1) tagged versions of the circuit. See Supplementary Information for complete model information.

Data Fig. 1a, b). Given the ultimate goal of implementation in an *in vivo* microenvironment characterized by variable growth conditions, we tested a range of incubation temperatures (36°C to 40°C) and perfusion rates (100 $\mu\text{m s}^{-1}$ to 200 $\mu\text{m s}^{-1}$), measuring an average period of 3 h across all conditions (Fig. 1d). These findings demonstrate that the SLC has the capacity to generate robust cycles of bacterial lysis in our microfluidic devices across a spectrum of environmental fluctuations that is likely to exist in an *in vivo* context.

The emergence of bacterial therapies in synthetic biology has accentuated the need for predictive modelling. This need stems from a bottleneck created by a difference in the timescales for bacterial cloning versus animal experiments; the circuits required for candidate therapies can be created much faster than they can be tested *in vivo*. Therefore, in order to quantitatively characterize the SLC concept before testing in animal models, we developed a computational model (Fig. 2a and Supplementary Information) to define an optimal strategy for subsequent testing in a lower-throughput animal model setting. We found that high production and degradation rates of the feedback-controlling proteins resulted in a wider domain of oscillatory dynamics in the parameter space (Fig. 2b). This model is consistent with our observations that oscillations in *S. Typhimurium* were more robust than in *Escherichia coli*, in which rates of protein production and degradation were previously found to be lower¹⁶ (Extended Data Fig. 1c and Supplementary Video 3). As the ability to manipulate circuit behaviour enhances the versatility of the system, we explored the tunability of the lysis period by adding an ssrA degradation tagging sequence on the LuxI protein. Consistent with model predictions, we observed an increased period and colony firing amplitude when tracking bacterial population dynamics (Fig. 2c, d and Extended Data Fig. 1d). The SLC thus enables tuning of the period and magnitude of delivery, which will be necessary for eventual application of this platform in the complex and fluctuating conditions present *in vivo*.

To incorporate a cytotoxic payload into the SLC strain, we added expression of Haemolysin E, encoded by *hlyE* of *E. coli*, which has been tested as a pore-forming anti-tumour toxin¹⁷. We initially confirmed the capability of the circuit to release intracellular contents by visualizing released sfGFP with a small microfluidic sink located beneath the growth chamber (Extended Data Fig. 2a–c). Then to visualize bacterial lysis and killing of cancer cells *in vitro* via HlyE, we engineered a microfluidic device so that cancer cells adhere inside a growth channel that is flanked by smaller bacterial growth chambers, which permits simultaneous single-cell visualization of bacterial lysis and cancer cell death (Extended Data Fig. 2d). After co-culturing human cervical

cancer HeLa cells with *S. Typhimurium* harbouring the SLC circuit, we observed HeLa cell death upon the onset of bacterial lysis, indicating efficient toxin release (Fig. 3a, b and Supplementary Videos 4 and 5). Complete cell death occurred in the growth channel within ~111 min of initial sfGFP fluorescence (Fig. 3c). Thus, the SLC bacteria were capable of releasing HlyE at levels necessary to kill tumour-derived cells.

We assessed the toxicity of released SLC or control bacterial contents in batch culture. As anticipated, we found that HeLa cells exposed to supernatant from a culture of the SLC bacteria bearing the *hlyE* module exhibited almost complete loss of viability (Fig. 3d), whereas the viability of HeLa cells exposed to supernatants of bacteria bearing the *hlyE* module without the SLC and equivalent dose of non-payload bearing SLC bacteria were only slightly affected (~15%). We concluded that bacterial lysis allowed for efficient HlyE release *in vitro* and that natural intracellular bacterial contents do not significantly affect HeLa cell viability. We further investigated the delivery characteristics of the SLC bacteria with *hlyE* by seeding variable amounts of circuit-harboring bacteria with HeLa cultures in well plates. We observed that the time to HeLa cell death following initial seeding increased with lower bacterial seeding volumes, presumably resulting from the extended time needed for bacteria to reach the quorum threshold (Fig. 3e and Supplementary Video 6). Initial seeding with a larger volume of bacteria resulted in increased firing rates which corresponded to shorter HlyE exposure times until cell death, consistent with a greater magnitude of lysis and payload release, although the cumulative toxicity threshold appears to be similar in all cases (Fig. 3f). On the basis of these observations, the seeding size of the bacterial population can be adjusted to determine the initial timing and release characteristics of the circuit.

We used a luciferase reporter to monitor bacterial population dynamics in grafted syngeneic colorectal tumours in mice. To minimize the extent of plasmid loss in the absence of antibiotic selection *in vivo*, we incorporated previously described stabilizing elements for plasmid retention and segregation into the SLC strain^{18–22}. Additionally, we placed both the payload and *luxCDABE* genes (the *in vivo* reporter module) under the *luxI* promoter as an indicator of *hlyE* production and quorum firing via bacterial luminescence (Fig. 1a). Using a subcutaneous model of colorectal cancer (MC26 cell line) in immunocompetent mice, we intratumorally injected a strain of SLC bacteria (SLC-hly). We observed pulsatile bacterial population dynamics within the tumour (Fig. 4a–c and Extended Data Fig. 3a, b) using *in vivo* imaging technology²³, consistent with the design and *in vitro* characterization (Fig. 2). The end luminescence intensity was on

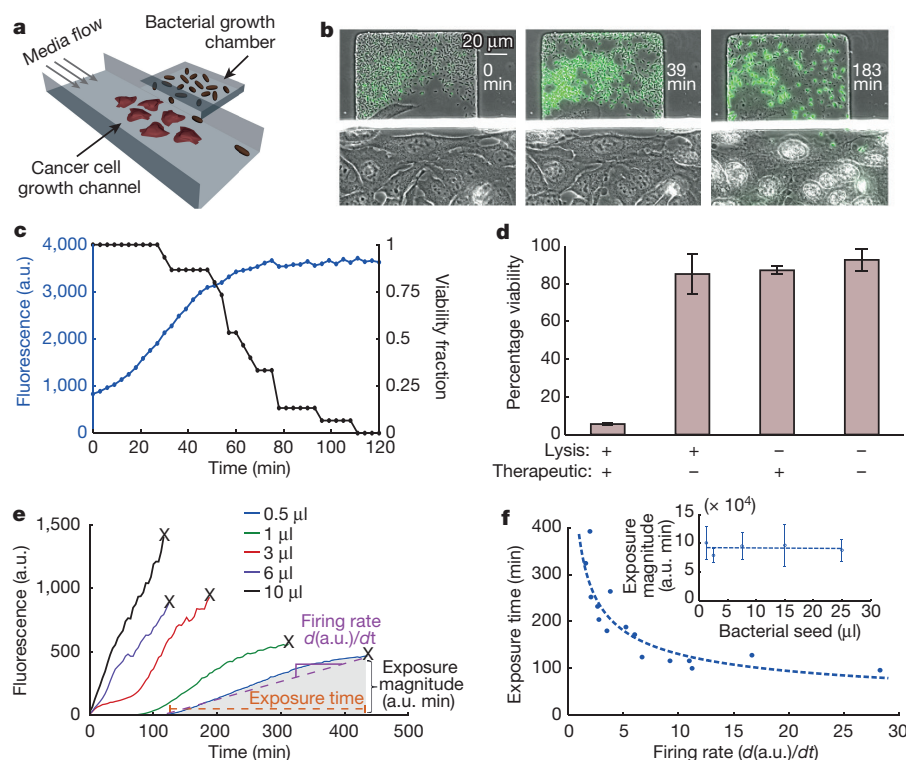


Figure 3 | In vitro co-culture. **a**, Schematic of the microfluidic co-culture with cancer cells and bacteria. Fluidic resistance was modified in this chip to achieve stable near-stagnant flow reduction to allow for cancer cell adherence and for diffusion of released therapeutic from the trap to the channel (see Supplementary Information). **b**, Frames from the co-culture time series sequentially visualizing *S. Typhimurium* (strain 3) firing, lysis and HeLa cell death. **c**, Fluorescent profile of the bacteria and HeLa cell viability fraction (number of live cells / number of dead cells in image frames) from **b** with time. **d**, Percentage viability of HeLa cells co-cultured with supernatant from *S. Typhimurium* culture harbouring the SLC + HlyE (strain 4), the SLC only (strain 5), constitutive *hlyE* only (strain 6), or no plasmid (strain 7). Error bars indicate ± 1 s.e. averaged over three measurements. **e**, Fluorescence profile of the SLC + HlyE (strain 4) co-cultured with HeLa cells at various initial seeding densities. The x symbols on the graph mark the point of complete HeLa cell death. **f**, The toxin exposure time, measured from the initial presence of fluorescence to HeLa cell death, as a function of the sfGFP production rate (see example in **e**). Although the time to death depends on seeding, the total magnitude of exposure remains conserved (inset). Error bars indicate ± 1 s.e. for three measurements. See Supplementary Information for ELH1301 host strain information.

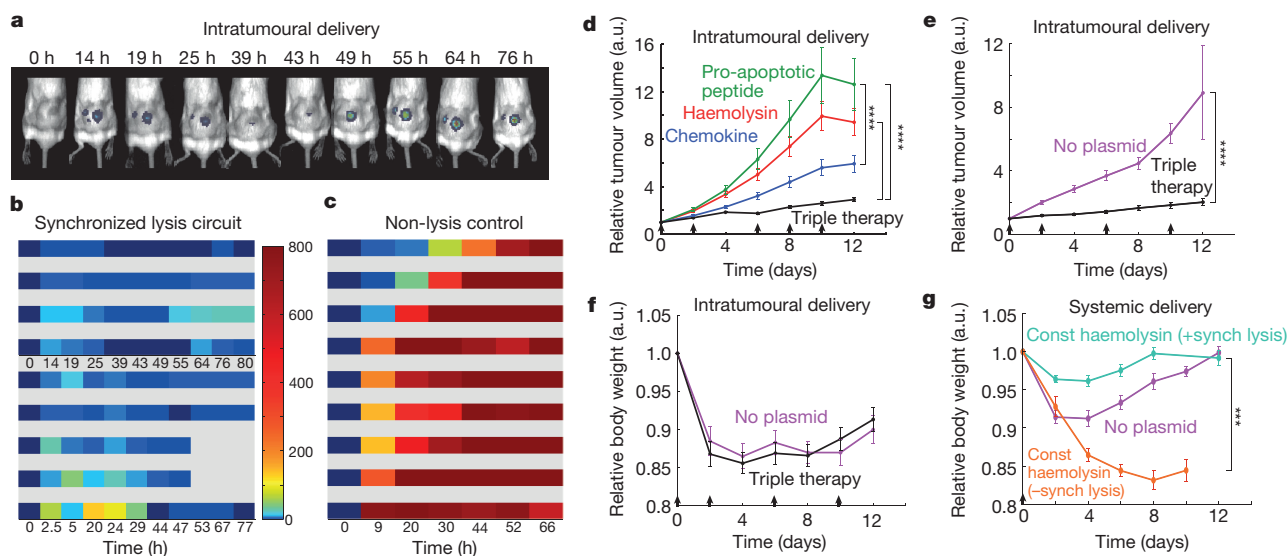


Figure 4 | In vivo bacterial dynamics, effect on tumours and tolerability in a subcutaneous tumour model. **a**, *In vivo* imaging over time of a mouse bearing two hind flank tumours injected once with the stabilized SLC-hly strain (strain 8). **b**, Single tumour density map trajectories of bacterial luminescence (relative to luminescence at 0 h) for the SLC-hly strain (strain 8). Data for each axis represents separate experiments. **c**, Single tumour density map trajectories of bacterial luminescence for the genomically integrated constitutively luminescent strain (strain 9). Intratumoural injection resulted in over 35-fold higher post-injection luminescence compared to intravenous injection (Extended Data Fig. 3d). **d**, Average relative tumour volume over time for subcutaneous tumour bearing mice injected with SLC-hly (red, strain 10), SLC-cdd (green, strain 14), SLC-ccl21 (blue, strain 15), and all together (SLC-3) (black). Bacteria were injected intratumorally on days 0, 2, 6, 8, and 10 (black arrows) ($***P < 0.0001$, two-way ANOVA with Bonferroni post-test, $n = 14$ –17 tumours, error bars show s.e.). **e**, Average relative tumour volume over

time for mice with subcutaneous tumours injected with the SLC-3 strains (black, strains 10, 14 and 15) and the no-plasmid control (magenta, strain 7). Bacteria were injected intratumorally on days 0, 2, 6, and 10 (black arrows) ($***P < 0.0001$, two-way ANOVA with Bonferroni post-test, $n = 18$ –19 tumours, error bars represent s.e.). **f**, Average relative body weight over time for mice with subcutaneous tumours injected with the SLC-3 strains (black, strain 10, 14, and 15) and the no-plasmid control (magenta, strain 7). Bacteria were injected intratumorally on days 0, 2, 6, and 10 (black arrows) ($n = 10$ mice for both cases, error bars represent s.e.). **g**, Average relative body weight over time for subcutaneous tumour-bearing mice with a single intravenous injection of the SLC + constitutive hlyE (turquoise, $n = 9$ mice, strain 11), a non-SLC strain with constitutive hlyE (orange, $n = 5$ mice, strain 12), or the no-plasmid control strain (magenta, $n = 9$ mice, strain 7) ($***P < 0.001$, two-way ANOVA with Bonferroni post-test, error bars represent s.e.).

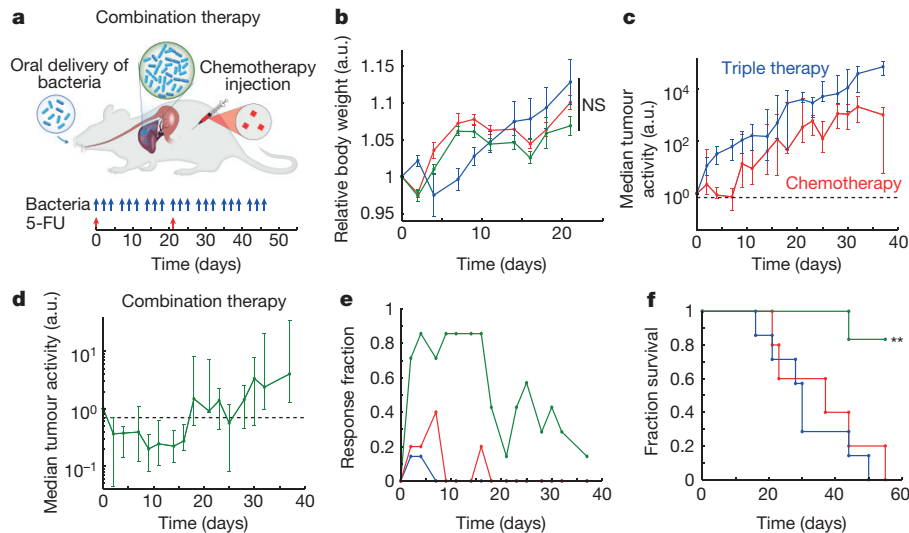


Figure 5 | *In vivo* testing in an experimental model of colorectal metastases in the liver via oral delivery of bacteria. **a**, Schematic of the experimental syngeneic transplantation model of hepatic colorectal metastases in a mouse, with the dosing schedule of either engineered bacteria (SLC-3) or a common cytotoxic chemotherapeutic, the antimetabolite 5-FU. The SLC-3 strains were delivered orally. 5-FU was delivered via intraperitoneal injection. **b**, Relative body weight over time for the mice with hepatic colorectal metastases fed with the SLC-3 strains (blue), injected with 5-FU chemotherapy (red), or a combination of the

two (green). Error bars indicate ± 1 s.e. for 5–7 mice. **c**, Median relative tumour activity, measured via tumour cell luminescence using *in vivo* imaging, for the chemotherapy and SLC-3 cases from **b**. **d**, Median relative tumour activity for the combination therapy case from **b**. Error bars for **c** and **d** indicate the interquartile ranges for 5–7 mice. The dashed line marks relative tumour activity of 0.70. **e**, Fraction of mice from the cases in **b** which respond with 30% reduction of tumour activity over time. **f**, Fraction survival over time for the mice in **b** (** $P < 0.01$, log rank test; $n = 5–7$ mice).

average ~ 300 -fold lower than the constitutive control strain, indicating a significant decrease in bacterial population levels within the tumour (Extended Data Fig. 3c).

Given the ability to engineer bacterial population dynamics in tumour grafts, we leveraged the versatility of the SLC bacteria as a delivery system to compare different classes of previously developed payloads. In addition to the haemolysin strain that was characterized in microfluidic devices, we created two additional SLC strains expressing genes to activate a host immune response (via T-cell and dendritic cell recruitment, using mouse CCL21) or trigger tumour cell apoptosis (using the cell death domain of Bit1 fused to the tumour-penetrating peptide iRGD, or CDD-iRGD)^{24,25}. Upon intratumoral injection, the immune recruitment strain elicited the strongest effect on tumour growth when compared to the haemolysin or apoptotic strains (Fig. 4d). We observed that an equal mixture of the three strains generated a stronger response than any single strain (Fig. 4d and Extended Data Fig. 3e–g), and on this basis we elected to pursue the ‘triple-strain’ dose for further testing in order to minimize animal usage. In a side-by-side comparison, we observed that the tumour response to SLC triple-strain (SLC-3) injections was significantly larger than the response to unmodified bacteria (Fig. 4e). Upon necropsy, histopathological analysis of remnant tumours was performed for mice treated with the SLC-3 strains, chemotherapy or unmodified bacteria. In mice treated with SLC-3 and non-circuit bacterial strains, robust staining of bacteria was observed by anti-*Salmonella* antibodies, showing localization of *Salmonella* within tumours. TUNEL staining indicated higher levels of apoptosis and cell death in SLC-3 treated tumours (Extended Data Fig. 4).

As a first step towards monitoring the effect of bacterial injections on the host, we compared how the triple-strain system affected body weight when administered intratumorally and intravenously, as the administration route affects bacterial localization (Extended Data Fig. 3d). We found that treatment with the SLC strains generated the same weight change as unmodified bacteria when administered intratumorally (Fig. 4f). However, intravenous administration of the SLC conferred a greater health benefit on the basis of observations that SLC strains producing constitutive therapy were better tolerated

than unmodified bacteria or non-SLC strains producing constitutive therapy (Fig. 4g). Although further targeted studies are required to systematically explore the effect of these bacteria on host health, these preliminary experiments suggest that the SLC design can reduce the burden of bacterial injections.

To explore a proof-of-principle for the application of our circuit in the context of *in vivo* tumours, we examined the efficacy of our system in an experimental syngeneic transplantation model of colorectal metastases within the liver. We had previously established that oral delivery of these bacterial strains led to safe and efficient colonization of hepatic colorectal metastases (see Methods), and that mice tolerated repeated dosing without overt adverse effects (Fig. 5a, b)²². In the context of bacteria-based therapeutic candidates, previous studies have shown that anaerobic bacteria can occupy avascular tumour compartments where chemotherapy is thought to be ineffective due to poor drug delivery¹¹. Thus a synergistic effect may arise when bacteria are used to deliver drugs to the necrotic core of a tumour, while standard chemotherapy is used for the vascularized regions^{11,26}. Inspired by this paradigm, we tested the combination of SLC-3 bacteria with a common clinical chemotherapy of 5-fluorouracil (5-FU). Tumours exhibited similar growth trajectories in response to repeated oral delivery of either the bacterial therapy alone, or two i.v. doses of 5-FU on day 0 and day 21 (Fig. 5c). In contrast, combination of these two applications led to a marked decrease in tumour activity over a period of 18 days, followed by a return to growth (Fig. 5d). During the initial 18-day period, a large fraction of the tumours was scored as eliciting at least a 30% reduction in tumour activity (Fig. 5e). The overall response led to roughly a 50% increase in the mean survival time for animals harbouring incurable colorectal metastases (Fig. 5f). Improvements may arise from strategies for long-term circuit stability or the utilization of additional therapeutic cargo.

The synchronized lysis circuit exemplifies a methodology for leveraging the tools of synthetic biology to exploit the ability of certain bacteria to colonize disease sites. In contrast to most drug delivery strategies, the synchronized lysis paradigm does not require pre-loading of a drug or the engineering of additional secretion machinery. In addition, it has the potential to decrease the likelihood of a systemic

inflammatory response through population control; as the bacterial colony is pruned after each oscillatory lysis event, the design could mitigate an undesirable host response. The circuit may enable new bacterial drug delivery strategies through modulation of the frequency and amplitude of the population cycles over time. Given recent insights into how host metabolism and circadian function are affected by the population dynamics of the gut microbiota, cyclical population control may be a prospective strategy to prevent host disturbances resulting from aberrant oscillations of gut microbes^{27,28}. Such engineering strategies may allow for the development of therapeutic communities within *in vivo* environments in which population dynamics are driven by interacting viruses, bacteria and host immune cells²⁹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 13 June 2016.

Published online 20 July 2016.

1. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nature Rev. Genet.* **13**, 260–270 (2012).
2. Xuan, C. *et al.* Microbial dysbiosis is associated with human breast cancer. *PLoS One* **9**, e83744 (2014).
3. Fischbach, M. A., Bluestone, J. A. & Lim, W. A. Cell-based therapeutics: the next pillar of medicine. *Sci. Transl. Med.* **5**, 179ps7 (2013).
4. Pawelek, J. M., Low, K. B. & Bermudes, D. Tumor-targeted *Salmonella* as a novel anticancer vector. *Cancer Res.* **57**, 4537–4544 (1997).
5. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic biology moving into the clinic. *Science* **333**, 1248–1252 (2011).
6. Weber, W. & Fussenegger, M. Emerging biomedical applications of synthetic biology. *Nature Rev. Genet.* **13**, 21–35 (2011).
7. Baban, C. K., Cronin, M., O'Hanlon, D., O'Sullivan, G. C. & Tangney, M. Bacteria as vectors for gene therapy of cancer. *Bioeng. Bugs* **1**, 385–394 (2010).
8. Hopton Cann, S. A., van Netten, J. P. & van Netten, C. Dr William Coley and tumour regression: a place in history or in the future. *Postgrad. Med. J.* **79**, 672–680 (2003).
9. Davila, M. L. *et al.* Efficacy and toxicity management of 19–28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci. Transl. Med.* **6**, 224ra25 (2014).
10. Garrett, W. S. Cancer and the microbiota. *Science* **348**, 80–86 (2015).
11. Dang, L. H., Bettgeowda, C., Huso, D. L., Kinzler, K. W. & Vogelstein, B. Combination bacteriolytic therapy for the treatment of experimental tumors. *Proc. Natl Acad. Sci. USA* **98**, 15155–15160 (2001).
12. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
13. Prindle, A. *et al.* A sensing array of radically coupled genetic 'biopixels'. *Nature* **481**, 39–44 (2011).
14. Young, K. D. & Young, R. Lytic action of cloned φ X174 gene E. *J. Virol.* **44**, 993–1002 (1982).
15. Marguet, P., Tanouchi, Y., Spitz, E., Smith, C. & You, L. Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology. *PLoS One* **5**, e11909 (2010).
16. Prindle, A. *et al.* Genetic circuits in *Salmonella typhimurium*. *ACS Synth. Biol.* **1**, 458–464 (2012).
17. Ryan, R. M. *et al.* Bacterial delivery of a novel cytolysin to hypoxic areas of solid tumors. *Gene Ther.* **16**, 329–339 (2009).
18. Gerdes, K. The *parB* (*hok/sok*) locus of plasmid R1: a general purpose plasmid stabilization system. *Nature Biotechnol.* **6**, 1402–1405 (1988).
19. Wood, T., Kuhn, R. & Peretti, S. Enhanced plasmid stability through post-segregational killing of plasmid-free cells. *Biotechnol. Tech.* **4**, 39–44 (1990).
20. Derman, A. I. *et al.* Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.* **73**, 534–552 (2009).
21. Danino, T., Lo, J., Prindle, A., Hasty, J. & Bhatia, S. N. *In vivo* gene expression dynamics of tumor-targeted bacteria. *ACS Synth. Biol.* **1**, 465–470 (2012).
22. Danino, T. *et al.* Programmable probiotics for detection of cancer in urine. *Science Transl. Med.* **7**, 289ra84 (2015).
23. Danino, T., Prindle, A., Hasty, J. & Bhatia, S. Measuring growth and gene expression dynamics of tumor-targeted *S. Typhimurium* bacteria. *JoVE* e50540 (2013).
24. Chen, R. *et al.* Application of a proapoptotic peptide to intratumorally spreading cancer therapy. *Cancer Res.* **73**, 1352–1361 (2013).
25. Loeffler, M., Le'Negrate, G., Krajewska, M. & Reed, J. C. *Salmonella typhimurium* engineered to produce CCL21 inhibit tumor growth. *Cancer Immunol. Immunother.* **58**, 769–775 (2009).
26. Forbes, N. S. Engineering the perfect (bacterial) cancer therapy. *Nature Rev. Cancer* **10**, 785–794 (2010).
27. Leone, V. *et al.* Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. *Cell Host Microbe* **17**, 681–689 (2015).
28. Thaiss, C. A., Levy, M. & Elinav, E. Chronobiomics: the biological clock as a new principle in host–microbial interactions. *PLoS Pathog.* **11**, e1005113 (2015).
29. Cheong, I. *et al.* A bacterial protein enhances the release and efficacy of liposomal cancer drugs. *Science* **314**, 1308–1311 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements The UCSD team was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM069811) and the San Diego Center for Systems Biology (P50 GM085764). The MIT team was supported by a Koch Institute Support Grant (P30-CA14051) from the National Cancer Institute (Swanson Biotechnology Center), a Core Center Grant (P30-ES002109) from the National Institute of Environmental Health Sciences, the Ludwig Center for Molecular Oncology at MIT and an Amar G. Bose Research Grant. T.D. was supported by the Misrock Postdoctoral fellowship and the NIH Pathway to Independence Award NIH (K99 CA197649-01). A.P. was supported by the Department of Defense National Defense Science and Engineering Graduate Fellowship and holds a Simons Foundation Fellowship of the Helen Hay Whitney Foundation and a Career Award at the Scientific Interface from the Burroughs Wellcome Fund. S.N.B. is an HHMI Investigator. We would like to thank R. Johnson for help with constructing microfluidic devices, H. Fleming for help with editing the manuscript, and H. Ding of The Barbara K. Ostrom (1978) Bioinformatics and Computing Facility in the Swanson Biotechnology Center for help with the statistical tests carried out in this study. We would also like to thank L. You for providing the lysis gene used in this study.

Author Contributions M.O.D. and J.H. designed the synchronized lysis circuit. M.O.D., J.S., L.S.T. and J.H. developed the computational model. M.O.D. and A.P. built and tested the bacterial circuit in microfluidics and performed the co-culture experiments. M.O.D. and E.J. collected the viability data, and M.O.D. and J.H. analysed the bacterial circuit data. T.D., M.S., K.A., and E.A. designed and performed the *in vivo* experiments. M.O.D., T.D., A.P., S.N.B., and J.H. analysed the animal data, and wrote and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H. (jhasty@eng.ucsd.edu).

Reviewer Information Nature thanks R. Solé, B. Vogelstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Strains and plasmids. Our circuit strains were cultured in LB media with $50 \mu\text{g ml}^{-1}$ and $34 \mu\text{g ml}^{-1}$ of kanamycin and chloramphenicol respectively, along with 0.2% glucose, in a 37°C incubator. Mammalian cells (HeLa CCL-2 from ATCC, verified by third-party cell line authentication services using an STR multiplex system) were cultured in DMEM media supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin (CellGro 30-002-CI), placed inside a tissue culture incubator at 37°C maintained at 5% CO_2 . Plasmids were constructed using the CPEC method of cloning or using standard restriction digest/ligation cloning. The activator plasmid (Kan, ColE1) was used in previous work from our group, and the lysis plasmid was constructed by taking the lysis gene, *E*, from the ePop plasmid via PCR and cloning it into a vector (Chlor, p15A) under the control of the *luxI* promoter^{13,15}. The *hlyE* gene was obtained via PCR from the genomic DNA of MG1655, while mouse CCL21 and CDD-iRGD were synthesized. These genes were cloned into the lysis plasmid, under the control of either the *tac* or *luxI* promoters (Extended Data Fig. 5). Co-culturing was performed with HeLa cells and either motile or non-motile *S. Typhimurium*, SL1344 (Extended Data Table 1). For full strain and plasmid information, please refer to the Supplementary Information.

Microfluidics and microscopy. The microfluidic devices and experiment preparation protocols used in this study are similar to those previously reported from our group¹³. The bacteria growth chambers were $100 \times 100 \mu\text{m}$ in area and approximately $1.4 \mu\text{m}$ in height. For co-culture experiments on the chip, we first loaded a suspended culture of HeLa cells in the device media channels at very low flow rates, to allow for adherence, and then incubated the device in a tissue culture incubator for 0.5–2 days to allow for proliferation. On the day of the experiment, the device was transferred to the microscope and circuit-containing bacteria were loaded in the growth chambers before imaging. Acquisition of images was performed with a Nikon TI2 using a Photometrics CoolSnap cooled CCD camera. The scope and accessories were programmed using the Nikon Elements software.

Co-cultures for well plate experiments were performed in Falcon 96-well tissue culture plates. HeLa cells were allowed to adhere to the wells before the addition of bacteria and subsequent imaging under the microscope or measurement in a TECAN Infinite M200 Pro plate reader. For viability measurements using an MTT assay, there were two technical replicates per well. For fluorescence measurements of co-cultures with variable seeding densities of bacteria, there were three technical replicates per case.

Additional details on microfluidics and microscopy can be found in the Supplementary Information.

In vivo experiments. All animal work was approved by the committee on animal care (MIT, protocol 0414-022-17). The protocol requires animals to be euthanized when tumors reach 2 cm^3 , or under veterinary staff recommendation. The cell line (MC26-LucF, Tanabe laboratory, Massachusetts General Hospital) was obtained from, and authenticated by, the Tanabe laboratory, MGH. The cell line was tested several times to be mycoplasma-free before implantation in mice. Sample sizes for mice were determined by expected effect size to produce a power of 0.8–0.9. Mice were blindly randomized into various groups using a random number generator.

Subcutaneous tumour model. Animal experiments were performed on 6-week-old female BALB/c mice (Taconic Biosciences) with bilateral subcutaneous hind flank tumours from an implanted mouse colon cancer cell line. The concentration for implantation of the tumour cells was 10^8 cells per ml in DMEM (no phenol red). Cells were then implanted subcutaneously at a volume of $100 \mu\text{l}$ per flank, with each implant consisting of 10^7 cells. Tumours were typically grown to an average of 300 mm^3 before experiments.

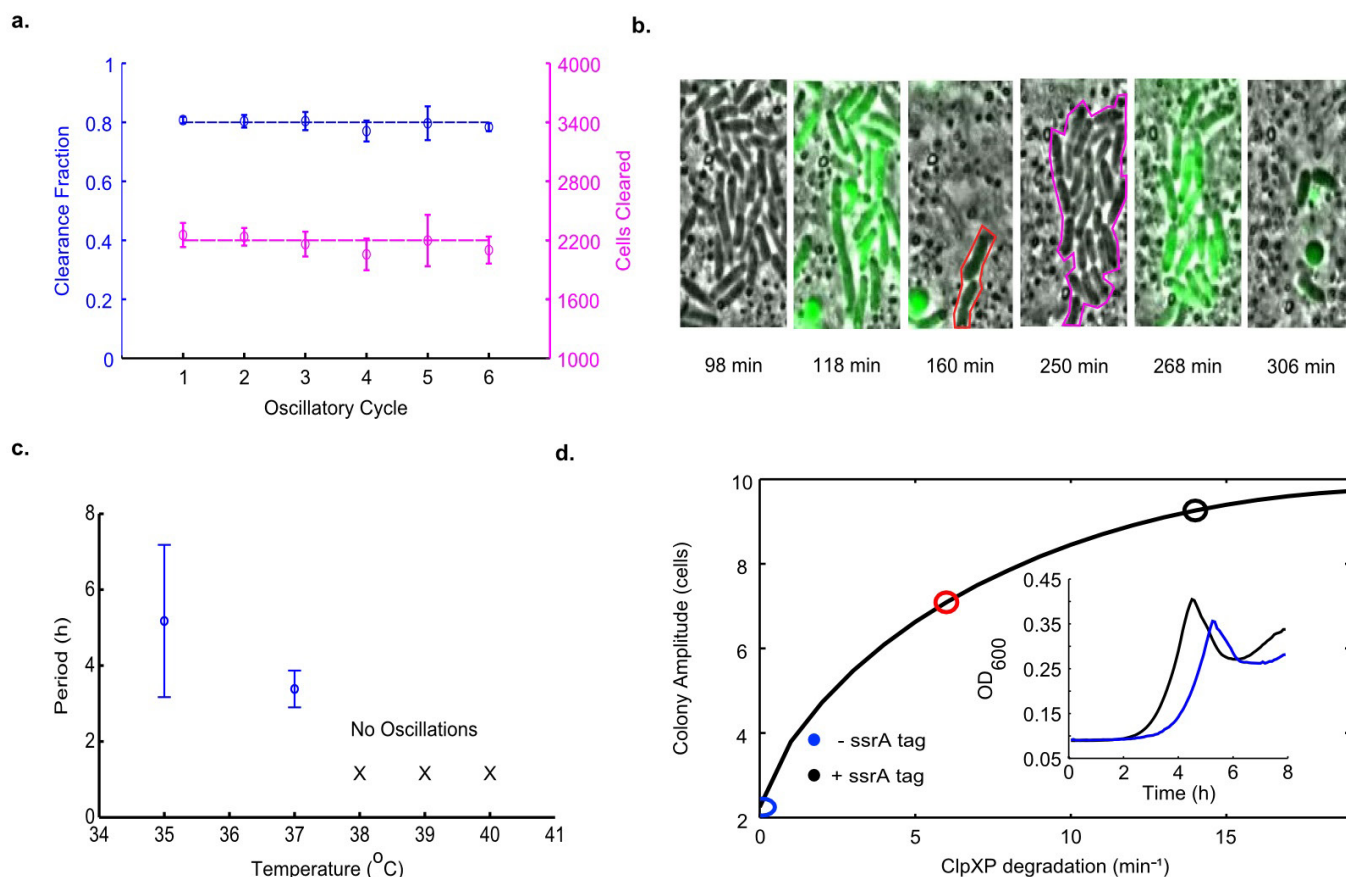
Experimental liver metastasis model. The experimental metastasis model was generated by injecting luciferase-producing mouse cancer cells into surgically externalized spleens of immunocompetent mice. Tumour cells seeded the liver during 90 s, after which the spleen was removed to prevent ectopic tumour growth³⁰. The MC26-LucF cell line was used (Tanabe Laboratory, MGH) and injected at 5×10^4 cells per $100 \mu\text{l}$ PBS into the spleens of female BALB/c mice at 6 weeks of age (Taconic Biosciences.). For the liver metastasis model, tumours were grown for 5–7 days to an average total tumour burden of 143 mm^3 before experiments.

Bacterial growth and administration. Bacterial strains were grown overnight in LB media containing appropriate antibiotics and 0.2% glucose as for the *in vitro* experiments. A $1:100 \times$ dilution in fresh media with antibiotics was started the day of injection and grown until an $\text{OD} < 0.1$ to prevent bacteria from reaching the quorum threshold (for SLC specifically). Bacteria were spun down and washed 2 to 3 times with sterile PBS before injection into mice. Intratumoural injections of bacteria were performed at a concentration of 5×10^7 cells per ml in PBS with a total volume of $10\text{--}20 \mu\text{l}$ injected per tumour, while intravenous injections were given at a total volume of $100 \mu\text{l}$. For the SLC-3 strains injection, this final volume was equally divided between the three strains at the indicated density. For liver metastasis experiments, bacteria were grown in LB media containing appropriate antibiotics and 0.2% glucose until they reached an OD of 0.05, after which they were concentrated to 10^9 to 5×10^9 bacteria per ml and delivered via oral gavage.

Post-administration monitoring for subcutaneous liver metastasis models. Luminescent signal was measured with the IVIS spectrum *in vivo* imaging system following bacterial injection. Measurements were compared relative to pre-injection values to follow dynamics. Subcutaneous tumour volume was quantified using calipers to measure the length, width, and height of each tumour throughout the imaging course ($V = L \times W \times H$). Volumes were compared to pre-injection values to follow physical tumour growth. Survival of mice was measured as the time from the beginning of the experiment up to the day when mice were moribund and euthanized. Survival for the experiment in Fig. 4f was measured with two biological replicates.

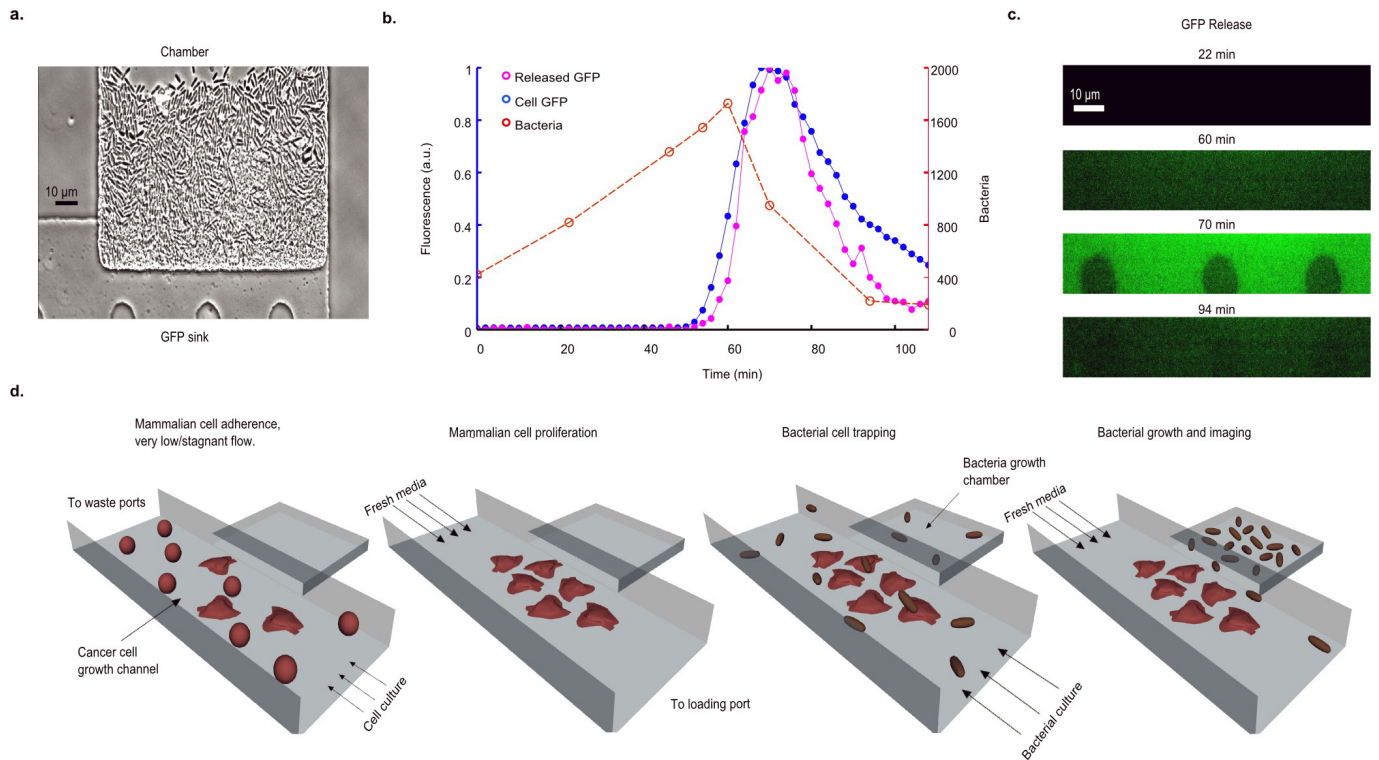
Statistical analysis. Statistical tests were calculated either in Excel (Student's *t*-test) or GraphPad Prism 5.0 (ANOVA with Bonferroni post-test, log-rank test). The details of the statistical tests carried out are indicated in the respective figure legends. Where data were approximately normally distributed, values were compared using either a Student's *t*-test or one-way ANOVA for single variable, or a two-way ANOVA for two variables. Mice were randomized in different groups before experiments.

30. Soares, K. C. *et al.* A preclinical murine model of hepatic metastases. *JoVE* e51677 (2014).



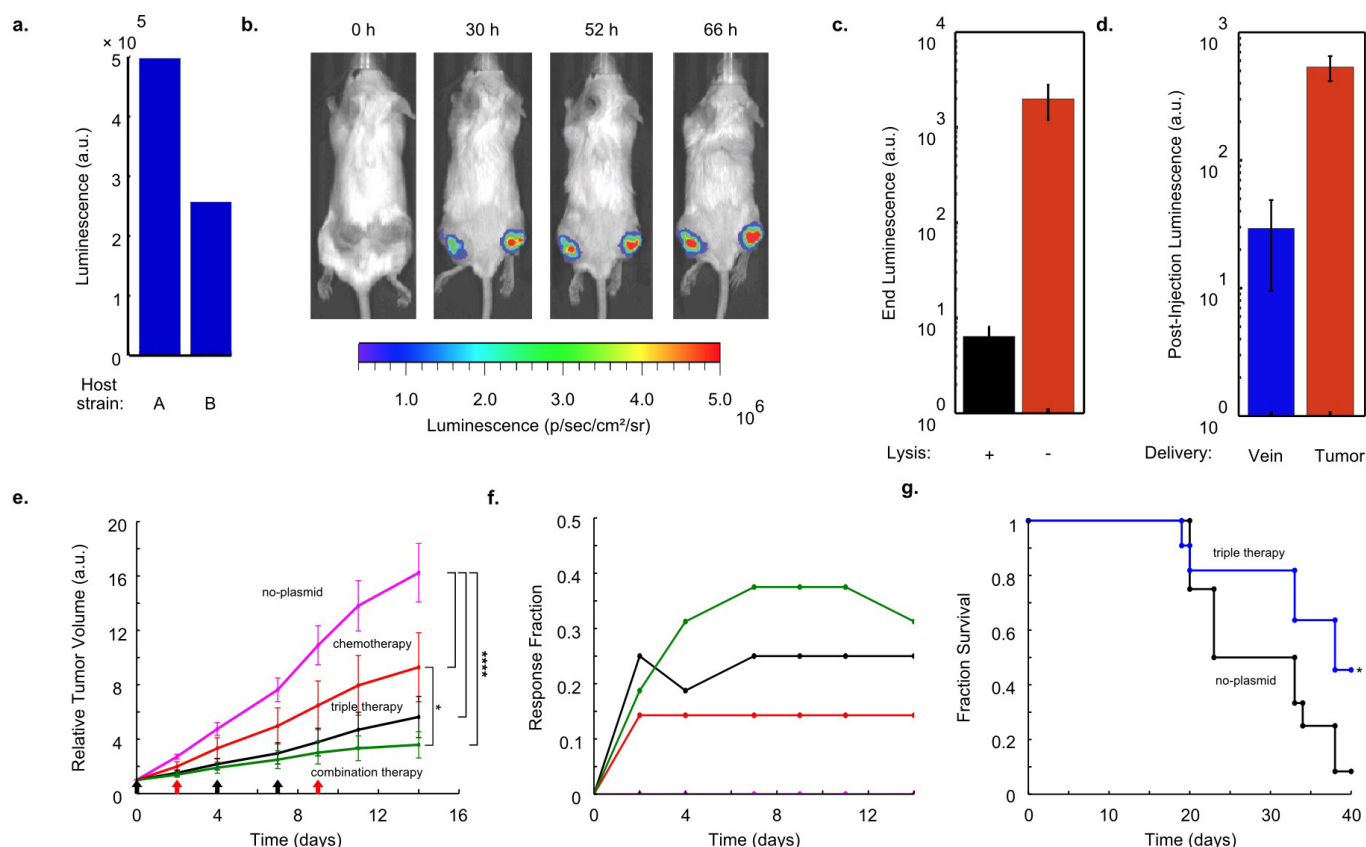
Extended Data Figure 1 | Various properties of the SLC. **a.** The fraction and number of bacterial cells cleared per consecutive oscillatory cycle in the growth chamber for a typical microfluidic experiment for *S. typhimurium*, including the effects of lysis and flow of cells outside of the trap (strain 1). **b.** Subset of time series images from the experiment in a showing a portion of the growth chamber where survivors of the initial lysis event (160 min frame, red outline) produce progeny (250 min frame, magenta outline) which are lysis sensitive. **c.** Period as a function of the

environmental temperature for *E. coli* (strain 13). The circuit does not oscillate for temperatures above 37 $^{\circ}\text{C}$ in *E. coli*. Error bars indicate ± 1 s.d. for 12–19 peaks. **d.** Colony amplitude at quorum firing for increasing degradation on the LuxI activator protein in the computational model. These simulation results are supported by batch well-plate experiments of the LuxI ssrA (black, strain 2) and non-ssrA (blue, strain 1) tagged versions of the circuit in *S. Typhimurium* (inset).



Extended Data Figure 2 | Investigating lysis-mediated intracellular release. **a**, A bacterial growth chamber with a $0.4\ \mu\text{m}$ high sink for sfGFP visualization after release. **b**, Number of bacteria (red), bacterial fluorescence (blue), sink fluorescence (pink) for a typical oscillatory

cycle (strain 1). **c**, Fluorescence time series images of the microfluidic sink from **b**. **d**, General procedure for performing bacterial and cancer cell co-culture experiments in a microfluidic device (also see Supplementary Information).

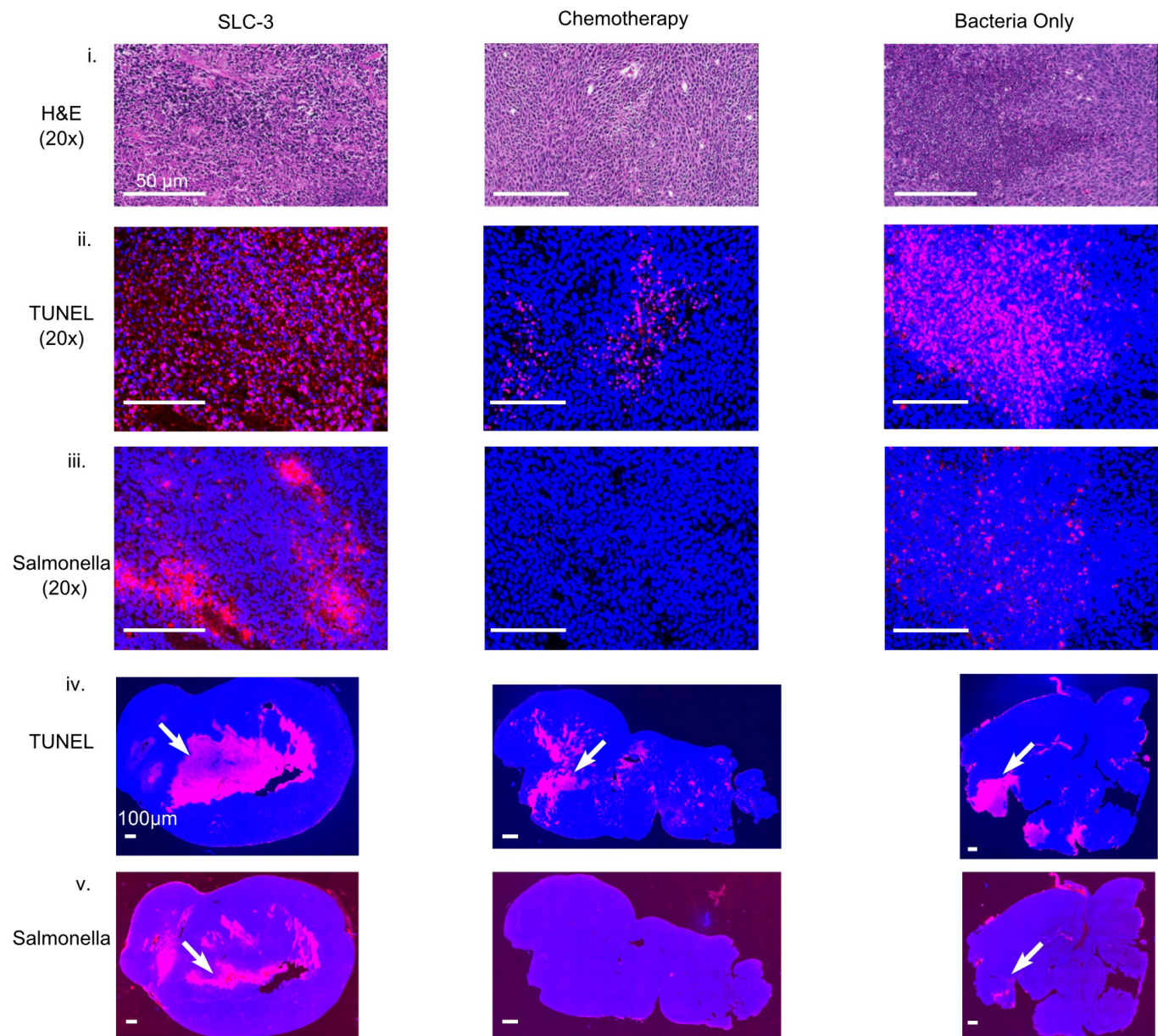


Extended Data Figure 3 | *In vivo* expression and therapy testing.

a. End-point *in vitro* luminescence intensity for SLC strains after ~20 h of growth. Host strains A and B are the host bacteria for strains 8 and 10. They are ELH1301 and ELH 430, respectively. Host A exhibits around twofold higher luminescence with the same circuit than host B. **b.** *In vivo* imaging over time of a mouse bearing subcutaneous tumours injected with a genomically integrated constitutively luminescent strain (strain 9). **c.** End-point *in vivo* bacterial luminescence of the SLC-hly strain and the constitutively luminescent strain from the experiments presented in Fig. 4. Error bars represent the s.e.m. bacterial luminescence from 9 tumours. **d.** Post-injection *in vivo* bacterial luminescence for the constitutively luminescent strain administered intravenously (vein) or intratumorally (tumour). Luminescence was measured ~20 h post-injection. Error bars

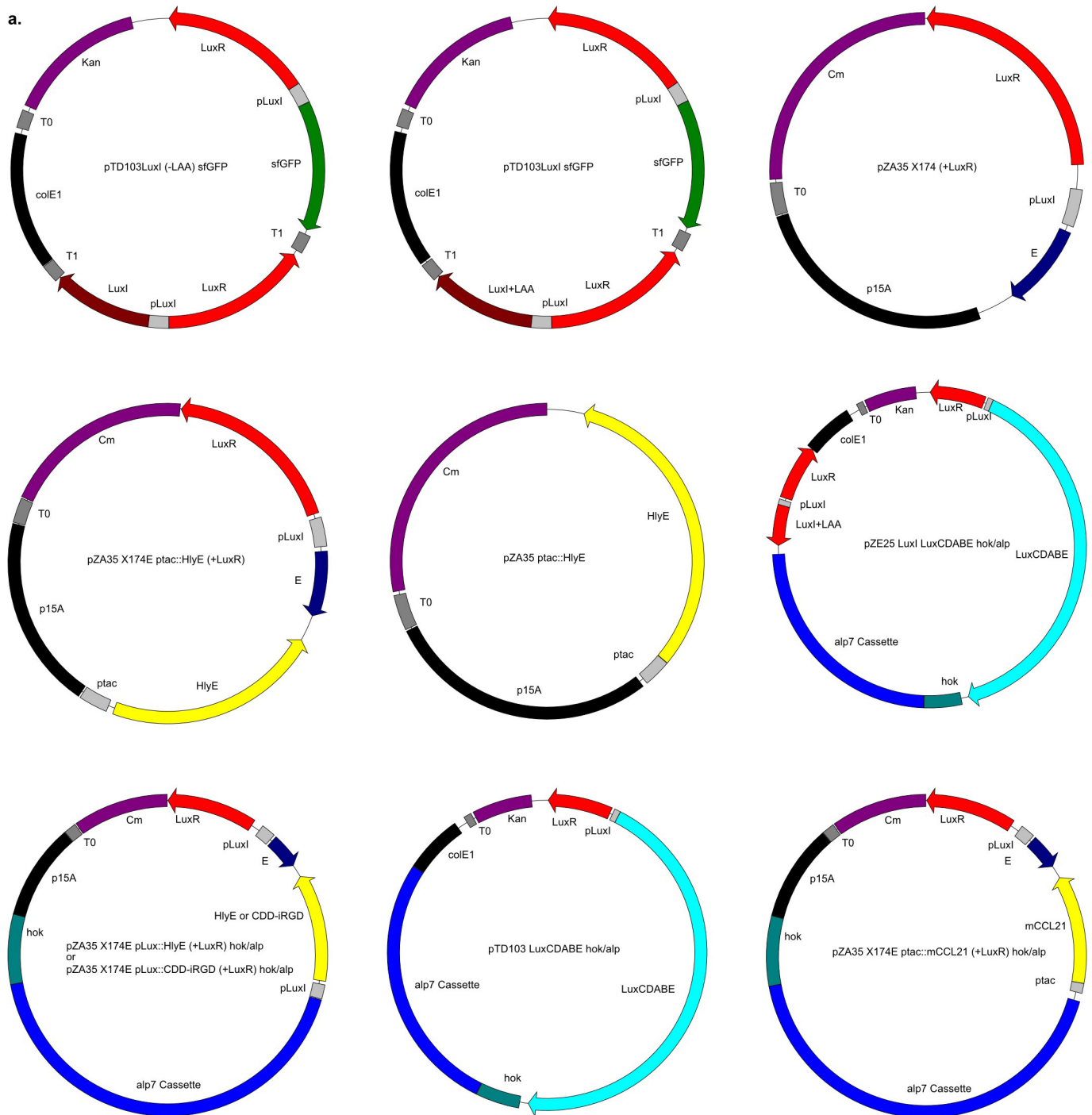
represent s.e.m. bacterial luminescence from 6 and 9 tumours for the intravenous and intratumoural cases, respectively. **e.** Average relative tumour volume over time for subcutaneous tumour bearing mice injected with the no-plasmid bacterium (strain 7), 5-FU chemotherapy, the SLC-3 strains, and the combination of SLC-3 with chemotherapy. Bacteria were injected intratumorally on days 0, 4, and 7 (black arrows), and chemotherapy was administered on days 2 and 9 (red arrows) (* $P < 0.05$, **** $P < 0.0001$, two-way ANOVA with Bonferroni post-test, $n = 12-16$ tumours, error bars represent s.e.). **f.** Fraction of mice from the cases in **e** which respond with 30% reduction of tumour volume over time. **g.** Fraction survival over time for mice with hepatic colorectal metastases fed with either the SLC-3 strains (blue) or the no-plasmid control (black) (* $P < 0.05$, log rank test; $n = 11-12$ mice).

a.

**Extended Data Figure 4 | Histological analysis of tumour sections.**

a, Histology of tumour sections taken from mice with different treatments 3 days post-administration. Haematoxylin and eosin staining for tissue sections intravenously injected with a combination of therapeutic bacteria (SLC-3), chemotherapy (5-FU), or a bacteria control with no therapeutic (strain 7) (i); TUNEL staining (red) in the same sections indicating cell apoptosis (ii); *Salmonella* immunohistochemistry (red) in the same sections confirming presence of bacteria in tumours (iii). Scale bars for (i–iii) denote 50 µm. TUNEL (iv) and *Salmonella* (v) staining (red) in

the entire tumour sections (examples indicated by arrows). Scale bars for iv and v denote 100 µm. DAPI staining (blue) was used to obtain a measure of live and dead cells in ii–iv. Histology slices ($n = 6$) from 20× images were compared across the groups and mean intensity of TUNEL staining, normalized by sample area, was demonstrated to be significantly higher for SLC-3 compared to the other two groups ($P < 0.0001$, one-way ANOVA), and not significantly different between the chemotherapy and bacteria-only cases.



Extended Data Figure 5 | The main plasmids used in this study. See Supplementary Information for more details.

Extended Data Table 1 | A list of strains and respective plasmids used in this study

Strain #	Strain Name	Host Bacterium	Plasmid(s)
1	MOD47	SL1344, M913	pTD103 luxI (-LAA) sfGFP + pZA35 X714E (+LuxR)
2	MOD46a	SL1344, M913	pTD103 luxI sfGFP + pZA35 X714E (+LuxR)
3	MOD67	SL1344, M913	pTD103 luxI (-LAA) sfGFP + pZA35 X714E (+LuxR) ptac::HlyE
4	MOD61	SL1344, ELH1301	pTD103 luxI sfGFP + pZA35 X714E (+LuxR) ptac::HlyE
5	MOD64	SL1344, ELH1301	pTD103 luxI sfGFP + pZA35 X714E (+LuxR)
6	MOD65	SL1344, ELH1301	pZA35 ptac::HlyE
7	ELH1301	SL1344, ELH1301	N/A
8	MOD105	SL1344, ELH430	pZE25 luxI luxCDABE hok/alp + pZA35 X714E (+LuxR) pLux::HlyE hok/alp
9	EcN-luxCDABE	Nissle 1917	N/A
10	MOD101	SL1344, ELH1301	pZE25 luxI luxCDABE hok/alp + pZA35 X714E (+LuxR) pLux::HlyE hok/alp
11	MOD102	SL1344, ELH1301	pZE25 luxI luxCDABE hok/alp + pZA35 X714E (+LuxR) ptac::HlyE hok/alp
12	MOD69	SL1344, ELH1301	pTD103 LuxCDABE hok/alp + pZA35 X714E (+LuxR) ptac::HlyE hok/alp
13	MOD29	JS006, BW25113	pTD103 luxI sfGFP + pZA35 X714E (+LuxR)
14	MOD110	SL1344, ELH1301	pZE25 luxI luxCDABE hok/alp + pZA35 X714E (+LuxR) pLux::CDD-iRGD hok/alp
15	MOD112	SL1344, ELH1301	pZE25 luxI luxCDABE hok/alp + pZA35 X714E (+LuxR) ptac::mCCL21 hok/alp

See Supplementary Information for more details.

CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis

Yoko Kojima¹, Jens-Peter Volkmer², Kelly McKenna², Mete Civelek³, Aldons Jake Lusis³, Clint L. Miller⁴, Daniel Drenzo¹, Vivek Nanda¹, Jianqin Ye¹, Andrew J. Connolly⁵, Eric E. Schadt⁶, Thomas Quertermous⁴, Paola Betancur², Lars Maegdefessel⁷, Ljubica Perisic Matic⁸, Ulf Hedin⁸, Irving L. Weissman² & Nicholas J. Leeper^{1,4}

Atherosclerosis is the disease process that underlies heart attack and stroke¹. Advanced lesions at risk of rupture are characterized by the pathological accumulation of diseased vascular cells and apoptotic cellular debris². Why these cells are not cleared remains unknown³. Here we show that atherogenesis is associated with upregulation of CD47, a key anti-phagocytic molecule that is known to render malignant cells resistant to programmed cell removal, or 'efferocytosis'^{4–7}. We find that administration of CD47-blocking antibodies reverses this defect in efferocytosis, normalizes the clearance of diseased vascular tissue, and ameliorates atherosclerosis in multiple mouse models. Mechanistic studies implicate the pro-atherosclerotic factor TNF- α as a fundamental driver of impaired programmed cell removal, explaining why this process is compromised in vascular disease. Similar to recent observations in cancer⁵, impaired efferocytosis appears to play a pathogenic role in cardiovascular disease, but is not a fixed defect and may represent a novel therapeutic target.

Each day the human body turns over more than 100 billion cells⁸. To prevent the inflammatory consequences associated with the accumulation of apoptotic debris⁹, these cells are rapidly and efficiently

cleared through a phagocytic process known as programmed cell removal, or 'efferocytosis'¹⁰. Programmed cell removal is mediated by macrophages detecting phagocytic 'eat me' signals on the target cell surface, and can be countermanded by cell-surface expression of anti-phagocytic 'don't eat me' signals such as expression of CD47 (ref. 6). Whereas programmed cell removal is highly conserved across almost all physiological conditions and in all tissues, it appears to be significantly impaired in atherosclerotic cardiovascular disease², the leading cause of death worldwide¹¹. Atherosclerosis is characterized by the accumulation of diseased macrophages and vascular smooth muscle cells (SMCs), which not only encroach on the lumen of the associated vessel but may also undergo programmed cell death^{1,12}. The impaired clearance of these diseased cells by lesional macrophages is thought to explain why these cells are frequently observed in the atherosclerotic necrotic core, and may potentiate vascular inflammation and risk for eventual plaque rupture^{3,13,14}. However, the mechanism underlying this defect has not yet been identified.

We recently found that the key anti-phagocytic molecule, CD47, is paradoxically upregulated by a variety of cancers^{5,7,15}. This renders malignant cells resistant to classic immune surveillance machinery

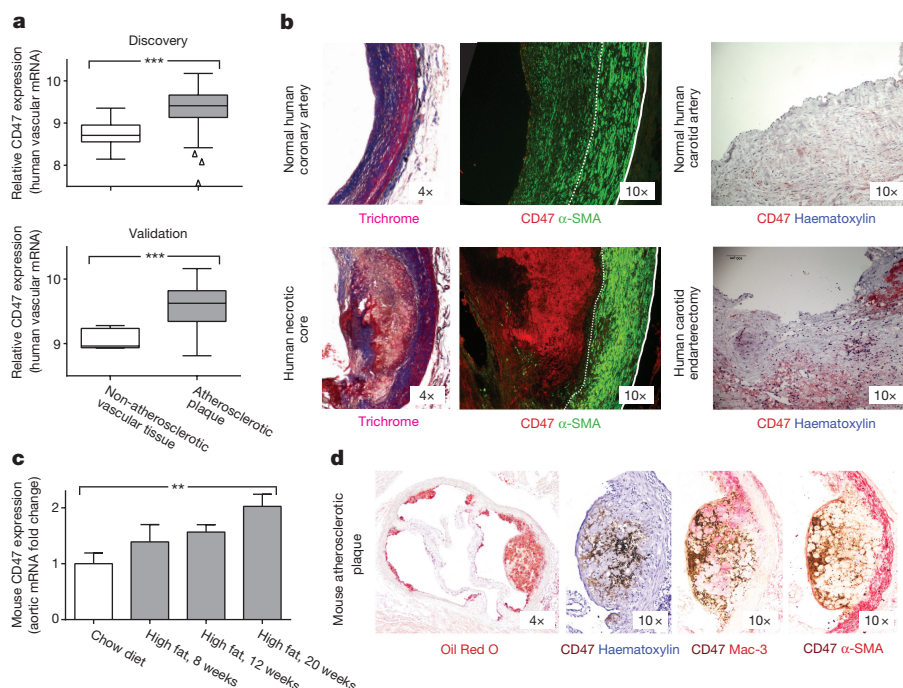


Figure 1 | CD47 is upregulated in atherosclerosis.

a, Microarray expression profiling in two carotid endarterectomy cohorts reveals that CD47 expression is significantly increased in human atherosclerotic plaque, relative to non-diseased vascular tissue (data displayed as Tukey box plots, $n = 182$ subjects). **b**, Immunostaining identifies intense CD47 upregulation within the necrotic core of human atherosclerotic coronary artery lesions (left) and carotid plaques (right). **c**, TaqMan mRNA analysis confirms that vascular CD47 expression progressively increases in a mouse model of atherosclerosis (*apoE*^{-/-} mice fed high-fat diet, grey), relative to control animals (C57BL/6 mice fed chow, white, $n = 4$ mice per time point). **d**, Immunohistochemistry staining with a biotin-labelled antibody (brown) reveals that CD47 expression co-localizes with apoptotic tissue in murine atherosclerotic plaque. *** $P < 0.001$, ** P for trend < 0.03 . Error bars represent the standard error of the mean (s.e.m.).

¹Department of Surgery, Division of Vascular Surgery, Stanford University School of Medicine, Stanford, California 94305, USA. ²Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. ³Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA. ⁴Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. ⁵Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA. ⁶Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York 10029, USA. ⁷Department of Medicine, Karolinska Institute, Stockholm, Sweden. ⁸Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden.

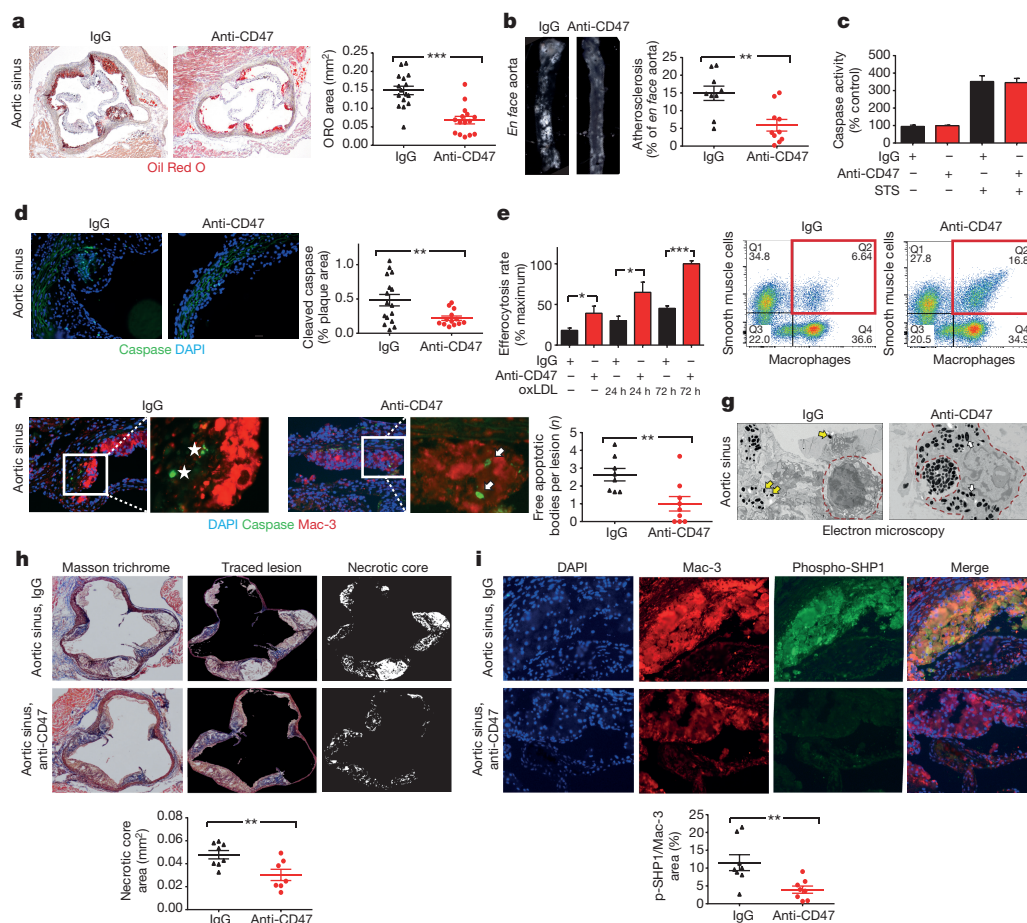


Figure 2 | Inhibition of CD47 stimulates efferocytosis and prevents atherosclerosis. **a**, Compared to mice treated with control antibodies (IgG, $n = 16$), mice treated with inhibitory anti-CD47 antibodies ($n = 15$) develop significantly smaller atherosclerotic plaques, as measured by Oil Red O (ORO) content in the aortic sinus. **b**, Total aortic atherosclerosis content is also reduced. **c**, **d**, Inhibition of CD47 signalling does not alter the rate of programmed cell death *in vitro* (**c**), but does reduce the accumulation of apoptotic bodies *in vivo* (**d**). **e**, Anti-CD47 antibody promotes efferocytosis of vascular cells at baseline and after exposure to pro-atherosclerotic lipids. Representative FACS phagocytosis plots for lipid-loaded (72 h) SMCs shown on the right (all assays repeated in triplicate). **f**, *In vivo* anti-CD47 antibody reduces the number of 'free' apoptotic bodies not associated with phagocytic macrophages, potentially

indicative of increased efferocytosis (stars indicate 'free' apoptotic bodies, arrows indicate 'not-free' apoptotic bodies). **g**, Electron microscopy confirms that mice treated with anti-CD47 antibodies display features of enhanced intraplaque efferocytosis, including an increased prevalence of macrophages which had ingested multiple apoptotic bodies (white arrows) and a reduced burden of 'free' apoptotic bodies (yellow arrows). **h**, Mice treated with anti-CD47 antibodies develop smaller necrotic cores than mice treated with IgG. **i**, Anti-CD47 antibody inhibits phosphorylation of lesional SHP1, a key anti-phagocytic effector molecule known to signal downstream of CD47. STS, staurosporine. Comparisons made by two-tailed *t*-tests. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Error bars represent s.e.m.

such as the tumoricidal macrophage, and is now recognized as a fundamental driver of tumour growth. To determine if dysregulated CD47 may also contribute to atherogenesis, we evaluated its expression in two independent human vascular tissue biobanks^{16,17}. We found that CD47 is consistently upregulated in human atherosclerotic plaque compared to non-atherosclerotic vascular tissue (Fig. 1a), and in subjects with symptomatic cerebrovascular disease (stroke or transient ischaemic attack) compared to those with stable asymptomatic lesions (Extended Data Fig. 1a). Because some efferocytosis molecules are known to undergo post-translational modification¹⁸, we also performed immunofluorescence and immunohistochemical staining of human coronary and carotid arteries, which confirmed that CD47 is progressively upregulated during atherogenesis and appears to localize intensely to the necrotic core (Fig. 1b, Extended Data Fig. 1b–g). Similar findings were observed in mouse models of atherosclerosis and other publically available microarray data sets (Fig. 1c, d, Extended Data Fig. 2). Together, these data suggest that pathologic upregulation of 'don't eat me' molecules may explain why phagocytosis is impaired within the human atherosclerotic plaque, which may promote lesion expansion over time.

To determine if this defect could be exploited as a translational target for cardiovascular disease, we treated a cohort of atheroprone animals (apolipoprotein-E-deficient (*apoE*^{−/−}) mice implanted with angiotensin-II-infusing minipumps¹⁹) with an inhibitory antibody directed against CD47 (Extended Data Fig. 3a)¹⁵. Compared to IgG control, anti-CD47 antibody treatment was associated with a marked reduction in atherosclerosis, both in the aortic sinus and *en face* in the aorta (Fig. 2a, b, Extended Data Fig. 3b, c). Similar results were observed in several additional models, including models of chronic atherosclerosis, plaque vulnerability and in mice with established lesions, as would be encountered clinically (Extended Data Fig. 3d–h). Although anti-CD47 antibodies had no effect on apoptosis *in vitro* (Fig. 2c, Extended Data Fig. 4a, b), we observed significantly fewer apoptotic bodies in the lesions of anti-CD47-treated animals *in vivo* (Fig. 2d). To reconcile this discrepancy, we used an established *in vitro* phagocytosis assay, and found that anti-CD47 antibodies potently induced the clearance of diseased and apoptotic vascular SMCs and macrophages that had been exposed to oxidized phospholipids to simulate the atherosclerotic environment (Fig. 2e, Extended Data Fig. 4c–f). Similarly, the number of 'free' apoptotic

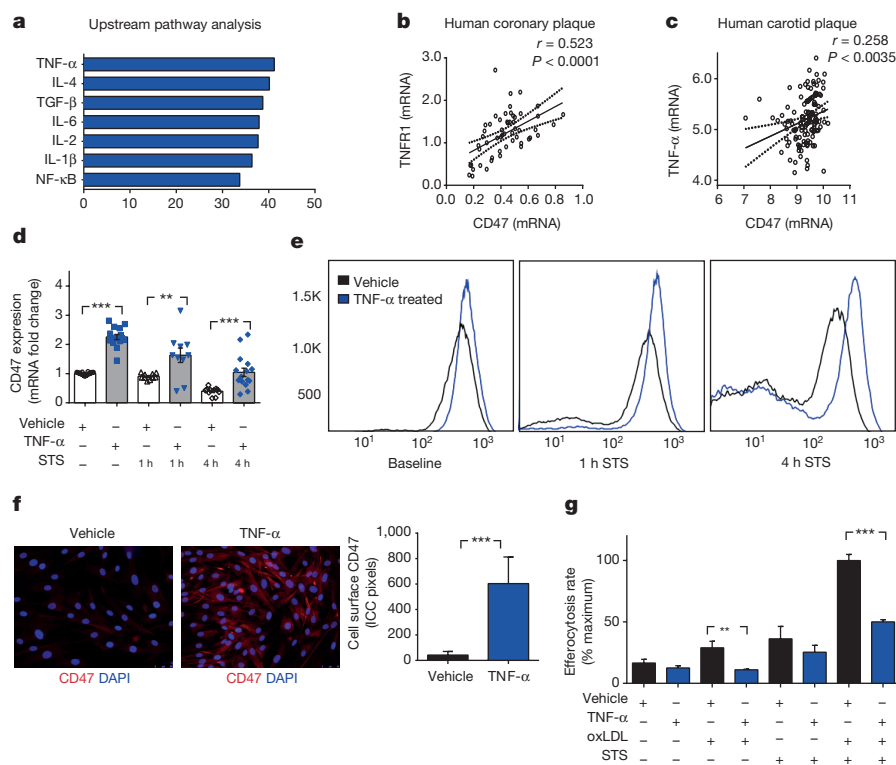


Figure 3 | The pro-atherosclerotic cytokine TNF- α induces CD47 expression and renders vascular cells resistant to phagocytic clearance.

a, Ingenuity Pathway Analysis identifies TNF- α as the regulator most likely to be upstream of genes that are co-expressed with CD47 in vascular tissue *ex vivo*. **b**, **c**, Co-expression studies confirm that CD47 is positively correlated with the canonical TNF- α receptor, TNFR1, in human coronary plaque (**b**) and TNF- α levels in human carotid plaque (**c**). The Pearson correlation coefficient was determined assuming a Gaussian distribution and *P* values were determined using a two-tailed test shown with the 95% confidence band of the best fit line. **d**, *In vitro*, TNF- α treatment

significantly increases the basal expression of CD47 in vascular SMCs, and blunts the decrease expected to occur during apoptosis. **e**, **f**, Flow cytometry (**e**) and fluorescent microscopy (**f**) confirm that TNF- α increases the cell-surface expression of CD47 on vascular cells at baseline and during programmed cell death. ICC, immunocytochemistry. **g**, *In vitro* efferocytosis assays indicate that TNF- α treatment renders vascular SMCs resistant to programmed cell clearance under a variety of pro-atherosclerotic conditions. Comparisons made by two-tailed *t*-tests. ****P* < 0.001, ***P* < 0.01. Error bars represent s.e.m.

bodies not associated with an intraplaque macrophage (indicative of failed efferocytosis) was reduced after anti-CD47 antibody treatment *in vivo*, as suggested by co-localization studies (Fig. 2f, Extended Data Fig. 5a) and electron microscopy (Fig. 2g, Extended Data Fig. 5b). Ultimately, these animals accumulated less apoptotic debris and developed lesions with smaller necrotic cores (Fig. 2h, Extended Data Fig. 5c). From a mechanistic perspective, anti-CD47 antibody therapy was associated with a marked suppression of intraplaque SHP1 phosphorylation, confirming interruption of the anti-phagocytic signalling axis downstream of SIRP α , the cognate anti-phagocytic receptor of CD47 (Fig. 2i, Extended Data Fig. 5d). These findings indicate that targeting CD47 can reduce atherosclerosis, and appears to do so by specifically reactivating efferocytosis within the lesion, without altering programmed cell death itself.

In addition to regulating phagocytosis, CD47 is also known to serve as a receptor for the vasoactive and nitric-oxide-regulating cytokine thrombospondin-1 (TSP1)²⁰. However, mice treated with anti-CD47 antibodies and IgG control had similar systemic blood pressures (Extended Data Fig. 6a) and rates of pulmonary nitric oxide elaboration (Griess reaction, Extended Data Fig. 6b), suggesting that anti-CD47 antibodies did not alter endothelial function *in vivo*. Further, anti-CD47 antibodies did not influence TSP1 signalling *in vitro*, having no modifying effect on TSP1-dependent MAPK signalling, eNOS phosphorylation, SMC proliferation or phagocytosis rates (Extended Data Fig. 6c–f). These data suggest that the anti-CD47 antibodies used in this study mediated anti-atherosclerotic effects independently of TSP1 signalling. Although full-dose anti-CD47 antibody therapy was again found to promote splenic erythrophagocytosis and compensatory

reticulocytosis (CD47 is a critical marker of self that is downregulated on ageing red blood cells)^{4,15}, this toxicity was self-limited and anaemia was not observed with chronic administration. Otherwise, this pro-efferocytic agent appeared to be well-tolerated, having no discernible effect on circulating leukocytes, lipid levels, or other metabolic parameters relevant to vascular disease (Extended Data Fig. 6g–u and Extended Data Table 1a).

To investigate why CD47 is upregulated in atherosclerosis, we used a bioinformatic approach in which we tested for genetic co-expression across panels of mouse and human vascular tissue. This yielded a list of genes that were significantly co-expressed with CD47 *in vivo* (Extended Data Fig. 7a). Pathway analyses of these genome-wide data sets identified 'inflammation mediated by chemokine and cytokine signaling pathway' as the top pathway associated with CD47 in vascular tissue (Extended Data Fig. 7b and Extended Data Table 1b, c), and specifically implicated TNF- α as the factor most likely to function as its upstream regulator (Fig. 3a). Subsequent correlation studies identified a strong positive association between CD47 and both the canonical TNF receptor, TNFR1 (Fig. 3b), as well as TNF- α itself (Fig. 3c), in human atherosclerotic vessel specimens. Similarly, CD47 expression levels were correlated with TNF- α levels in tissue from atherosclerotic mice and confirmatory human data sets (Extended Data Fig. 7c, d). Together, these informatics and co-expression studies implicate TNF- α —a proinflammatory cytokine known to be upregulated in atherosclerosis—in CD47-dependent vascular disease^{21,22}.

To investigate the causality of these associations, we next tested whether CD47 is directly downstream of TNF- α . We found that treatment of vascular SMCs with recombinant TNF- α led to a

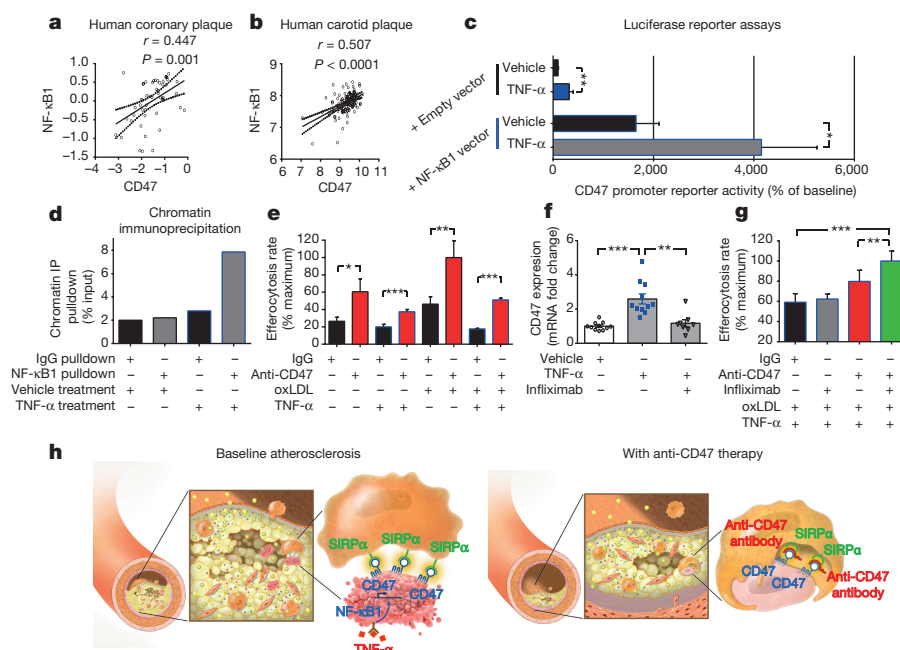


Figure 4 | TNF- α promotes CD47 expression via NF- κ B1 and is a translational cardiovascular target. **a, b**, Co-expression analyses confirm that NF- κ B1 is significantly correlated with CD47 expression in both human coronary (**a**) and carotid (**b**) atherosclerotic plaque. Pearson correlation coefficients were determined assuming a Gaussian distribution and *P* values were determined using a two-tailed test. **c**, Dual luciferase reporter assays reveal that *CD47* promoter activity is stimulated in cells treated with TNF- α (top), but that this effect is significantly enhanced in cells co-transfected with an NF- κ B1 expression vector. **d**, Chromatin immunoprecipitation studies confirm significant enrichment of NF- κ B1 protein on the *CD47* promoter in TNF- α -treated human coronary artery SMCs. **e**, *In vitro* efferocytosis assays reveal that anti-CD47 antibody

enhances the clearance of cells exposed to TNF- α , and that its pro-efferocytic capacity is enhanced under pro-atherosclerotic conditions. **f**, Pretreatment with the anti-TNF- α monoclonal antibody infliximab prevents the upregulation in *CD47* mRNA that normally occurs in SMCs exposed to TNF- α . **g**, Concomitant inhibition of CD47 and TNF- α using anti-CD47 antibodies and infliximab, respectively, produces synergistic benefit in the clearance of diseased vascular cells, as assessed by ANOVA. **h**, Putative mechanism explaining why efferocytosis is impaired in cardiovascular disease, and how inhibition of CD47-SIRP α signalling could represent a new therapeutic target. Comparisons made by two-tailed *t*-tests, unless otherwise specified. ****P* < 0.001, ***P* < 0.01, **P* < 0.05. Error bars represent s.e.m.

consistent upregulation of cellular CD47 expression, whereas no effect was observed with a variety of other common pro-atherosclerotic or proinflammatory insults (Extended Data Fig. 7e–h). Further, TNF- α blunted the progressive decrease in CD47 expression normally expected to occur during programmed cell death (Fig. 3d–f, Extended Data Fig. 7i–l). As a result of their higher levels of anti-phagocytic molecules, TNF- α -treated cells were less likely to be phagocytosed by macrophages, particularly when concomitantly exposed to oxidized low-density lipoproteins (oxLDL) and pro-apoptotic stimuli (Fig. 3g, Extended Data Fig. 7m). Because impaired efferocytosis is known to incite proinflammatory cytokine elaboration⁹, it is possible that a positive feedback loop underlies the co-localization of TNF- α , CD47, and uncleared pathological cells and apoptotic bodies within the atherosclerotic plaque²².

Analysis of the *CD47* promoter in vascular SMCs revealed a region of open chromatin predicted to contain binding sites for several of the NF- κ B transcription factors known to be downstream of TNFR1 (Extended Data Fig. 8a and Extended Data Table 1d). Among these, the classical proinflammatory factor NF- κ B1 (p50) was found to be positively correlated with CD47 expression in both human coronary and carotid plaques (Fig. 4a, b, Extended Data Fig. 8b). Luciferase reporter assays performed with a vector containing the *CD47* promoter revealed that TNF- α treatment stimulated basal CD47 expression, and that the effect was specifically enhanced when NF- κ B1 was simultaneously overexpressed in these cells (Fig. 4c, Extended Data Fig. 8c–e). Chromatin immunoprecipitation assays confirmed that NF- κ B1 binds to the *CD47* promoter *in vitro*, and that this occupancy is increased several fold upon treatment with TNF- α (Fig. 4d, Extended Data Fig. 8f).

From a translational perspective, we found that anti-CD47 antibodies were able to stimulate efferocytosis in TNF- α -treated cells, and that the

effect was most pronounced under dyslipidaemic, pro-atherosclerotic conditions (Fig. 4e). A modest incremental benefit was observed when anti-CD47 antibody therapy was combined with commercially available anti-TNF- α therapies, such as infliximab or etanercept, probably because of their inhibitory influence on CD47 expression in mouse and human tissue (Fig. 4f, g, Extended Data Fig. 9). These data are particularly provocative given the observation that patients prescribed TNF- α -inhibiting antibodies for inflammatory disorders such as lupus and rheumatoid arthritis appear to be protected from myocardial infarction^{21,23}.

The finding that CD47 expression is pathologically upregulated in both cancer and cardiovascular disease suggests a commonality between these two conditions. In leukaemogenesis, cancer stem cells out-compete normal haematopoietic stem cells, while countering signalling associated with programmed cell removal; viable myelodysplastic syndrome haematopoietic oligolineage progenitors express the phagocytic signal calreticulin, but not CD47, whereas acute myeloid leukaemia derived from myelodysplastic syndromes are positive for CD47 expression²⁴. Similar cellular processes in the vasculature may explain the recent observations that de-differentiated SMCs undergo clonal expansion within the atherosclerotic plaque^{25,26}. Furthermore, the top cardiovascular locus identified by genome-wide association studies surprisingly resides near an important tumour suppressor locus²⁷, which in turn regulates SMC efferocytosis¹⁷. Future studies will need to examine whether expansion of CD47^{hi} SMC clones contributes to atherosclerosis, and if their clearance can be accomplished without the induction of anaemia (for example, with a dose-escalation approach that appears to be safe in non-human primates²⁸ and is currently being pursued in first-in-human clinical trials²⁹).

Together, these data provide insights into why programmed cell removal is impaired in the atherosclerotic plaque, and how this may

promote lesion expansion. Our findings bolster the ‘inflammatory hypothesis’ of atherosclerosis³⁰, and specifically link cytokine signalling with anti-phagocytic signalling in vascular disease. Given the experimental success of pro-efferocytic therapies in the oncology field using antibodies that block the CD47 signal⁵, it is possible that these findings will provide a novel nonsurgical treatment of cardiovascular disease (mechanism shown in Fig. 4h).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 March; accepted 16 June 2016.

Published online 20 July; corrected online 3 August 2016

(see full-text HTML version for details).

- Libby, P., Ridker, P. M. & Hansson, G. K. Progress and challenges in translating the biology of atherosclerosis. *Nature* **473**, 317–325 (2011).
- Schrijvers, D. M., De Meyer, G. R., Kockx, M. M., Herman, A. G. & Martinet, W. Phagocytosis of apoptotic cells by macrophages is impaired in atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1256–1261 (2005).
- Thorp, E. & Tabas, I. Mechanisms and consequences of efferocytosis in advanced atherosclerosis. *J. Leukoc. Biol.* **86**, 1089–1095 (2009).
- Oldenborg, P. A. *et al.* Role of CD47 as a marker of self on red blood cells. *Science* **288**, 2051–2054 (2000).
- Chao, M. P., Majeti, R. & Weissman, I. L. Programmed cell removal: a new obstacle in the road to developing cancer. *Nat. Rev. Cancer* **12**, 58–67 (2011).
- Gardai, S. J. *et al.* Cell-surface calreticulin initiates clearance of viable or apoptotic cells through trans-activation of LRP on the phagocyte. *Cell* **123**, 321–334 (2005).
- Majeti, R. *et al.* CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells. *Cell* **138**, 286–299 (2009).
- Kinchen, J. M. & Ravichandran, K. S. Phagocytic signaling: you can touch, but you can't eat. *Curr. Biol.* **18**, R521–R524 (2008).
- Fadok, V. A. *et al.* Macrophages that have ingested apoptotic cells *in vitro* inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF- β , PGE₂, and PAF. *J. Clin. Invest.* **101**, 890–898 (1998).
- Henson, P. M., Bratton, D. L. & Fadok, V. A. Apoptotic cell removal. *Curr. Biol.* **11**, R795–R805 (2001).
- World Health Organization. The top ten causes of death. <http://www.who.int/mediacentre/factsheets/fs310/en/> (2014).
- Shankman, L. S. *et al.* KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat. Med.* **21**, 628–637 (2015).
- Tabas, I. Macrophage death and defective inflammation resolution in atherosclerosis. *Nat. Rev. Immunol.* **10**, 36–46 (2010).
- Thorp, E., Cui, D., Schrijvers, D. M., Kuriakose, G. & Tabas, I. MERTK receptor mutation reduces efferocytosis efficiency and promotes apoptotic cell accumulation and plaque necrosis in atherosclerotic lesions of apoE^{-/-} mice. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1421–1428 (2008).
- Willingham, S. B. *et al.* The CD47-signal regulatory protein alpha (SIRP α) interaction is a therapeutic target for human solid tumors. *Proc. Natl Acad. Sci. USA* **109**, 6662–6667 (2012).
- Perisic, L. *et al.* Profiling of atherosclerotic lesions by gene and tissue microarrays reveals PCSK6 as a novel protease in unstable carotid atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **33**, 2432–2443 (2013).
- Kojima, Y. *et al.* Cyclin-dependent kinase inhibitor 2B regulates efferocytosis and atherosclerosis. *J. Clin. Invest.* **124**, 1083–1097 (2014).
- Garbin, U. *et al.* Expansion of necrotic core and shedding of MERTK receptor in human carotid plaques: a role for oxidized polyunsaturated fatty acids? *Cardiovasc. Res.* **97**, 125–133 (2013).
- Daugherty, A., Manning, M. W. & Cassis, L. A. Angiotensin II promotes atherosclerotic lesions and aneurysms in apolipoprotein E-deficient mice. *J. Clin. Invest.* **105**, 1605–1612 (2000).
- Isenberg, J. S. *et al.* CD47 is necessary for inhibition of nitric oxide-stimulated vascular cell responses by thrombospondin-1. *J. Biol. Chem.* **281**, 26069–26080 (2006).
- Bäck, M. & Hansson, G. K. Anti-inflammatory therapies for atherosclerosis. *Nat. Rev. Cardiol.* **12**, 199–211 (2015).
- Hopkins, P. N. Molecular biology of atherosclerosis. *Physiol. Rev.* **93**, 1317–1542 (2013).
- Greenberg, J. D., Furer, V. & Farkouh, M. E. Cardiovascular safety of biologic therapies for the treatment of RA. *Nat. Rev. Rheumatol.* **8**, 13–21 (2011).
- Pang, W. W. *et al.* Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. *Proc. Natl Acad. Sci. USA* **110**, 3011–3016 (2013).
- Feil, S. *et al.* Transdifferentiation of vascular smooth muscle cells to macrophage-like cells during atherogenesis. *Circ. Res.* **115**, 662–667 (2014).
- Chung, I. M., Schwartz, S. M. & Murry, C. E. Clonal architecture of normal and atherosclerotic aorta: implications for atherogenesis and vascular development. *Am. J. Pathol.* **152**, 913–923 (1998).
- McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
- Liu, J. *et al.* Pre-clinical development of a humanized anti-CD47 antibody with anti-cancer therapeutic potential. *PLoS One* **10**, e0137345 (2015).
- ClinicalTrials.gov: NCT02216409.
- Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by the National Institutes of Health (R01HL12522401 and R01HL12337001 to N.J.L. and U01HL099999 to I.L.W.) and the Ludwig Center at Stanford. The authors wish to acknowledge J. Knowles for his critical review of the manuscript.

Author Contributions Y.K. designed and conducted most experiments including the mouse microsurgery, morphometric analyses, *in vitro* cell culture, TaqMan, western blot, ChIP and luciferase reporter experiments. V.N., D.D. and L.M. assisted with the electron microscopy studies and blinded histology. J.Y., J.P.V. and K.M. conducted the *in vitro* efferocytosis assays and FACS studies. L.P. and U.H. performed the carotid staining and BiKE biobank analyses. A.C. provided coronary samples and assisted with the histopathological studies. M.C., A.J.L., E.S., P.B., T.Q. and C.M. conducted the microarray, co-expression and *in silico* bioinformatic studies. Y.K., I.L.W. and N.J.L. conceived the study, analysed the data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.L. (nleeper@stanford.edu).

Reviewer Information *Nature* thanks I. Tabas and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. The experiments were not randomized. The investigators were not blinded to allocation during experiments, but were blinded during data analysis and interpretation. No statistical methods were used to predetermine sample size.

Human cardiovascular tissue. *Carotid endarterectomy samples.* In this study, a total of 182 human carotid endarterectomy samples and nonatherosclerotic control arteries were used. These include heterogeneous atherosclerotic plaque samples obtained from patients undergoing surgery for symptomatic (stroke or transient ischaemic attack) or asymptomatic (no history of cerebrovascular event) high-grade carotid stenosis (>50% NASCET criteria) as part of the Biobank of Karolinska Endarterectomies (BiKE). 15 nonatherosclerotic control arteries (iliac artery and aorta) were obtained from organ donors without any history of cardiovascular disease. Patients were consecutively enrolled in the study, with the first 127 constituting the discovery cohort (40 asymptomatic, 87 symptomatic) and the next 50 constituting the validation cohort (10 asymptomatic, 40 symptomatic). All samples were collected with informed consent from patients, organ donors or their guardians. The BiKE study was approved by the Ethical Committee of the Northern Stockholm. DNA and RNA was extracted from these specimens and analysed by Illumina 610w -QuadBead SNP-chips and the Affymetrix HG-U133 plus 2.0 microarrays (discovery cohort) or the Affymetrix HG-U133a Genechip arrays (validation cohort), as previously described^{16,17} and deposited in Gene Expression Omnibus (accession number GSE21545). Robust multi-array average (RMA) normalization was performed and processed gene expression data was returned in log₂-scale. The relative expression of each gene was determined by comparing the pixel intensity of the 11 probe pairs which correspond to each transcript to the 'normalization control set' specific to each array (see http://www.affymetrix.com/support/technical/technotes/expression_comparison_technote.pdf and http://www.affymetrix.com/support/technical/datasheets/hgu133arrays_datasheet.pdf). Student's *t*-test with correction for multiple comparisons according to the Sidak-Bonferroni method was used for statistical analyses of microarray data. Pearson's correlations were calculated to determine associations between expression of the gene of interest and other genes from microarrays. Publically available data from the Helsinki Carotid Endarterectomy Study (HeCES) were analysed as a second human validation cohort³¹. Additional publically available vascular and nonvascular microarray data deposited in GEO were also analysed (as indicated by GSE number in the corresponding figure legend), including studies of microdissected atherosclerotic plaques and samples taken from individuals treated with commercially available TNF- α inhibitors. A *P* value <0.05 was considered to indicate significance.

Coronary artery samples. In this study, a total of 114 human coronary artery samples were used. 51 atherosclerotic epicardial coronary artery segments were harvested from 22 orthotopic heart transplant donors, as previously described³². Additionally, 56 coronary artery segments were extracted from atherosclerotic intracoronary plaques in patients undergoing coronary atherectomy, as previously described³³. Briefly, longitudinal atherectomy was performed with the Silverhawk atherectomy catheter in arteries with flow-limiting stenosis after diagnostic angiography. RNA was isolated from both sets of specimens and hybridized to custom dye gene expression microarrays representing approximately 22,000 features^{32,33}. Briefly, the custom probe set was identified from data mining and curation of atherosclerosis and vascular cell-culture-based expression analyses and combined with the Agilent Human 1A and 1B arrays. In each microarray experiment, the expression levels of CD47 were determined by array-specific hybridization probe sets (~10 per transcript). Automated feature extraction software was used to filter out background or saturated spot intensities to eliminate inherent probe and spatial biases during the fluorescent detection. This software computes the log₂ ratios and filters features by *P* value. These values were normalized according to protocol for each array, using a locally weighted linear regression curve fit (LOWESS) to correct for dye biases. This normalization procedure results in array specific log₂ ratios, and provides accurate measures of relative gene expression differences (http://www.affymetrix.com/support/technical/datasheets/human_datasheet.pdf). Normalized data from candidate transcripts were then used to calculate a Pearson correlation coefficient *r* assuming a Gaussian distribution, and two-tailed *P* values were calculated for each correlation coefficient.

In addition to samples obtained for RNA analysis, an additional seven right coronary arteries were obtained from rapid autopsies from adult patients with a spectrum of coronary artery disease for histological analysis. The arteries were fixed with 4% paraformaldehyde (PFA) for several hours, and immersed in 30% sucrose at 4 °C overnight. The coronaries were then serially sectioned and segments of atheroma and relatively normal coronary artery were embedded in paraffin or OCT, and sectioned at 7- μ m thickness.

For immunohistochemical staining, paraffin slides were deparaffinized and rehydrated with xylene and alcohol gradients, and antigen retrieval was performed

in sodium citrate buffer (10 mM sodium citrate, 0.05% Tween 20, pH 6.0) using a pressure cooker. The sections were blocked with 5% goat serum, and then stained with primary antibodies (anti-CD47, Abcam B6H12.2, 2 μ g ml⁻¹) or mouse IgG1 kappa (eBioscience, 2 μ g ml⁻¹) at 4 °C overnight. The sections were washed in TBS, incubated with MACH4 kit (Biocare Medical) according to the manufacturer's instructions, and then detected using the Vulcan Fast Red Chromogen kit2 (Biocare Medical). Staining for SMC content was accomplished by washing sections in water and PBS, followed by incubation with anti-SM22 alpha antibodies (Abcam, ab14106, 1:300). The sections were washed with PBS, incubated with Alexa Fluor 488 goat anti-Rabbit (Life technologies, 1:250), washed, and mounted with Vectashield Mounting medium with DAPI (Vector Laboratories). Pictures were taken by Nikon digital camera mounted on an inverted fluorescence microscope. Serial sections were prepared as described above, and also stained with Masson's trichrome (Richard-Allan), or haematoxylin and eosin (Richard-Allan). For the staining of frozen sections, OCT was removed in water, and the sections were stained with Oil Red O (ORO, Sigma-Aldrich, O0625, 0.5%), haematoxylin and eosin (H&E, Richard-Allan), smooth muscle alpha-actin (Abcam, ab5694, 1:300), HMGB1 (Abcam, ab18256, 1:100), CD206 (Abcam, ab64693, 1:50), CD47 (Abcam B6H12.2 or Novus Biologicals, #NBP2-31106, 1:50) or mouse IgG1 kappa (eBioscience). Secondary antibodies included Alexa Fluor 594 goat anti-mouse (Life technologies, A11005, 1:300) and Alexa Fluor 488 goat anti-rabbit (Life technologies, A11034, 1:300). Antibody specificity was confirmed using isotype control and by preincubating the anti-CD47 antibodies with recombinant CD47 antigen (R&D Systems) in a 1:5 ratio for 16h at 4 °C before application. High resolution imaging of the carotid sections was performed as previously described³⁴.

Murine cardiovascular tissue. *Atherosclerosis models.* In the atherosclerosis studies described below, a total of 179 male *apoE*-deficient mice on the C57BL/6 background (*apoE*^{-/-}, Jackson Laboratory, catalogue #002052) were used.

In the main atherosclerosis intervention studies, 8-week-old mice were implanted with subcutaneous Alzet minipumps (model 2004, Alzet Osmotic Pumps) containing Angiotensin II (AngII, Sigma-Aldrich, 1000 ng kg⁻¹ min⁻¹) and initiated on a high fat Western diet (21% anhydrous milk fat, 19% casein and 0.15% cholesterol, Dyets no. 101511) for the ensuing 4 weeks, as previously described¹⁹. To determine the effect of CD47 signalling on vascular disease, mice were injected with either 200 μ g of the inhibitory anti-CD47 antibodies (MIAP410, BioXcell, *n* = 18) or IgG1 control (MOPC-21, BioXcell, *n* = 20) IP QOD, at the dose previously studied¹⁵. The antibody therapy was started one day before the pump implantation. Animals were observed daily, and in the case of premature sudden death, necropsy was performed to determine the cause of mortality. Blood pressure in conscious mice was measured at baseline (after standard acclimatization with a Visitech Systems Inc. machine), and weekly for the duration of the study. At 12 weeks of age, the mice were killed after an overnight fast, with serum and visceral organs (including the aortae) isolated and processed for analysis.

In the chronic atherosclerosis studies, male *apoE*-deficient mice were weaned and initiated on a high fat diet at 4 weeks of age and maintained on this for the subsequent 12 weeks (without any angiotensin infusion). Antibody injection was performed as described above while the high fat diet was given (*n* = 9 per condition), and the animals were killed at the age of 16 weeks.

In the established disease model, *apoE*-deficient mice were weaned onto a high fat diet at 4 weeks of age and continued on this for the ensuing 8 weeks (without any angiotensin infusion). At 12 weeks of age (after lesions had developed), mice were initiated on 200 μ g of anti-CD47 antibodies (*n* = 14) or IgG1 (*n* = 13) IP QOD for the ensuing 6 weeks, and killed at 18 weeks of age.

In the TNF- α inhibitor synergy studies 8-week-old *apoE*-deficient mice were implanted with AngII pumps and maintained on Western high-fat diet, and then randomized to one of four groups, including the: (1) IgG group (mouse IgG and human Fc, *n* = 10); (2) etanercept group (etanercept 0.2 mg per kg SQ weekly (Amgen) and mouse IgG IP QOD, *n* = 8); (3) CD47 antibody group (MIAP410 50 μ g IP QOD and human Fc SQ weekly, *n* = 16); or (4) combination group (MIAP410 and etanercept, *n* = 19). In this model, the antibody treatment was started the day before AngII pump implantation and delivered for 4 weeks, and the mice were euthanized at 12 weeks of age. Note that in this model, the anti-CD47 antibody dose was reduced by 75% to determine whether a lower dose of therapy could also affect atherogenesis.

In the short-term intervention model, 8-week-old *apoE*-deficient mice were implanted with AngII pumps and maintained on Western high-fat diet without antibody therapy for the ensuing 23 days. Beginning at day 23, the mice received SQ injections of either: (1) IgG daily, *n* = 10; (2) etanercept (0.8 mg per kg at day 23, *n* = 6); (3) anti-CD47 antibody (200 μ g of MIAP410 daily, *n* = 11); or (4) combination therapy, *n* = 11 between days 23 and 27. This cohort of mice was killed after only 5 days of antibody treatment (at day 28) and was used to evaluate the effect of antibody therapy in established atherosclerotic plaques of identical size.

To evaluate the effect of antibody therapy on atherosclerotic plaque vulnerability, we used the recently described 'tandem stenosis' model³⁵. At 6 weeks of age, *apoE*-deficient mice were initiated on high-fat diet and maintained on this for ensuing 6 weeks. At 12 weeks of age, tandem stenoses were introduced in a manner shown to reproducibly alter shear stress and induce plaque rupture, as previously described³⁵. Briefly, the mice were anaesthetized by isoflurane inhalation and an incision was made to allow dissection of the right common carotid artery from the circumferential connective tissues. Serial stenosis with a 150 µm outer diameter were then introduced 1 mm and 4 mm from the carotid bifurcation. The stenosis diameter was obtained by placing a 6-0 suture around the carotid artery together with a 150 µm needle that was tied to it and later removed. Antibody therapy was started the day before the surgery and continued thereafter. Mice were killed 7 weeks after the surgery and intraplaque haemorrhage was quantified within the processed tissue sections.

In addition to the mice treated with anti-CD47 antibody or control antibody, a separate cohort of 12 male *apoE*^{-/-} mice were fed a high-fat diet for either 8, 12 or 20 weeks, but were not treated with antibody nor implanted with osmotic minipumps. These mice were used to determine aortic gene expression changes during atherogenesis, using the RNA analysis methods described below. In these experiments, comparison was made to control C57BL/6 mice fed standard chow diet for 24 weeks.

Finally, a cohort of 3 *apoE*^{-/-} mice were implanted AngII osmotic pump and fed a high-fat diet for 4 weeks (but not treated with antibody) and then injected with biotin-labelled anti-CD47 antibodies 24 and 6 h before being killed, to determine where the therapeutic antibody accumulates *in vivo*. Nonatherosclerotic C57BL/6 and CD47-deficient mice (*Cd47*^{-/-}, Jackson Laboratory, catalogue #003173) were also injected with the biotin-labelled antibodies, and served as controls. All animals were analysed as described below.

Vascular tissue preparation, immunohistochemistry and atherosclerotic lesion quantification. Aortic atherosclerosis lesion area was determined as described previously¹⁷. Briefly, the arterial tree was perfused with PBS and then fixed with 4% PFA. The heart and the full-length of the aorta-to-iliac bifurcation was exposed and dissected carefully from any surrounding tissues. Thoracic aortas were then opened along the ventral midline and dissected free of the animal and pinned out flat, intimal side up, onto black wax. Aortic images were captured with a digital camera mounted on a Nikon stereomicroscope and analysed using Adobe Photoshop CS5 software. The percentage of lesion area was calculated as total lesion area divided by total surface area. The atherosclerotic lesions within the aortic valve area (aortic sinus) were analysed as described previously¹⁷. The samples were perfused with PBS, fixed with 4% PFA, embedded in OCT, and sectioned at 7-µm thickness. Four sections at 100-µm intervals were collected from each mouse and stained with ORO, Masson's trichrome, haematoxylin and eosin, smooth muscle α -actin (SMA, Abcam, ab5694, 1:300), Mac-3 (BD Sciences, BD 550292, 1:100), CD-3 (Abcam, ab5690, 1:150), and Ly-6G (BD Sciences, BD 551459, 1:300). Atherosclerosis burden was quantified from the luminal aspect of the blood vessel through the plaque to the internal elastic lamina (that is, lipid in the neointima was quantified). Necrotic core size was quantified by calculating the area of the lesion which was acellular on Masson's trichrome staining, as previously described^{17,36}. Plaque haemorrhage was quantified by determining the presence or absence of red blood cells (TER-119, Santa Cruz Biotechnology) within the plaque, as previously described³⁷. Subsequent immunohistochemical studies were quantified from the luminal aspect of the blood vessel through the plaque to the external elastic lamina (to assess changes which also involved the tunica media). To detect the localization of injected biotin-labelled anti-CD47 antibodies, the avidin-biotin complex technique was used. Frozen sections of aortic sinus were prepared from the mice injected with biotin-labelled anti-CD47 antibodies, as described above. Endogenous peroxidase activity was blocked by incubation with 0.3% hydrogen peroxide for 30 min, and the sections were washed with water and PBS, followed by blocking with 5% goat serum for 30 min. Biotin was detected using Vectastain ABC kit and DAB substrate kit per protocol (Vector laboratories). Corresponding aortic sinus sections from the mouse without biotin antibody injection were used as negative controls, as were *Cd47*^{-/-} mice which had been injected as above. For immunofluorescent staining of these samples, sections were blocked with 5% goat serum for 30 min, then incubated with Streptavidin-Alexa Fluor 546 conjugate (Life Technologies, 1:300) and Mac-3 (BD, 1:100) for 1 h, followed by Alexa Fluor 488 donkey anti-rat (Life Technologies, 1:300). *In vivo* apoptosis was assessed by staining for TUNEL positivity with the Cell Death Detection Kit (Roche), per protocol, and confirmed with cleaved caspase 3 (Cell Signaling #9661, 1:200) staining followed by Alexa Fluor 488 goat anti-rabbit (Life Technologies, 1:250). The cleaved-caspase-3-positive area was measured and quantified using Adobe Photoshop, and the percentage of positive area was calculated as total caspase-3-positive area divided by total atherosclerotic plaque area measured by

ORO staining in the serial sections. To calculate the *in vivo* phagocytic index, we performed double staining of cleaved caspase 3 (detected with Alexa Fluor 488 goat anti-rabbit antibody) and Mac-3 (detected with Alexa Fluor 594 goat anti-rat antibody (Life technologies, 1:250)). The number of free apoptotic cells not associated with a macrophage (indicated by a star (Fig. 2f)) was manually assessed in a blinded fashion, and compared to apoptotic cells associated with a macrophage (indicated by an arrow), as previously described². For phospho-SHP1 staining, the sections were stained with phospho-SHP1 antibodies (Abcam, ab131500, 1:50) and Mac-3 followed by Alexa Fluor. The phospho-SHP1-positive area was normalized to Mac-3-positive area. All lesion areas and indices were measured and quantified using Adobe Photoshop by a blinded observer. Samples collected from several tissue beds were also snap-frozen in liquid nitrogen for subsequent mRNA and protein expression analysis, as described below.

Vascular tissue electron microscopy. Electron microscopy was performed in the Stanford University Cell Sciences Imaging Facility, as previously described¹⁷. Briefly, samples were fixed and processed using standard histologic techniques then imaged using a JEOL JEM-1400 transmission electron microscope. Ingested apoptotic bodies (indicated by white arrow (Fig. 2g)), free apoptotic bodies (indicated by yellow arrow), and apoptotic bodies undergoing secondary necrosis (indicated by red arrow) were qualitatively assessed in a blinded manner, as previously described¹⁷.

Serum and plasma analysis. Serum chemistry, lipid, complete blood count, and differential analyses were performed by the Stanford Animal Diagnostic Laboratory, as previously described¹⁷. In brief, blood samples were collected by cardiac puncture after an overnight fast. Automated haematology was performed on the Sysmex XT-2000iV analyser system. Blood smears were prepared for all full complete blood count samples and reviewed by a medical technologist. Chemistry analysis was performed on the Siemens Dimension Xpand analyser, and included analyses of renal function, electrolyte levels, liver function tests, fasting glucose levels and fasting lipid panels. A medical technologist performed all testing, including dilutions and repeat tests as indicated, and reviewed all data. Serum insulin levels were measured by ELISA kit according to the manufacturer's instruction (EMD Millipore).

Griess reaction. The activity of nitric oxide synthase was evaluated using a modified Griess assay. Lung samples were collected from mice and snap-frozen by liquid nitrogen before homogenization in PBS. The nitrate and nitrite levels were measured by Ultrasensitive Colorimetric Assay for nitric oxide synthase (Oxford Biomedical Research) according to the manufacturer's instructions, and standardized by the protein amount.

Hybrid Mouse Diversity Panel. The Hybrid Mouse Diversity Panel (HMDP), which includes a quantitative analysis of 109 classical and recombinant inbred mouse strains³⁸, was used to identify factors associated with vascular CD47 expression, *in vivo*. Briefly, whole aorta from the arch to the mid-abdomen was snap-frozen at the time of death and total RNA was isolated using the RNeasy kit (Qiagen), as described³⁹. Genome-wide expression profiles were determined by hybridization to Affymetrix HT-MG_430 p.m. microarrays on a subset of female mice from 104 strains ($n = 2$ aorta per strain). Quantification of plasma cytokines was carried out in a multiplexed immune-capture microbead system (Milliplex Mouse Cytokine/Chemokine Magnetic Bead Panel MCYTOMAG-70K, EMD Millipore) as per the manufacturer's instructions. Cytokines profiled were: G-CSF, GM-CSF, IFN γ , IL-1 α , IL-1 β , IL-2, IL-4, IL-6, IL-7, IL-10, IL-12 (p40), IL-12 (p70), IL-13, IL-15, IP-10, KC, MCP-1, MIP-1 α , MIP-1 β , M-CSF, MIP-2, MIG, RANTES, TNF- α . Plasma insulin was measured using the mouse insulin ELISA kit (80-INSMS-E01, Alpco) as per the manufacturer's instructions. Pearson's correlations were generated to calculate transcript-transcript and transcript-trait correlations. Using these methods, the genes and plasma cytokines that were significantly associated with aortic CD47 expression levels were identified.

***In silico* bioinformatics methods. Pathway analysis.** Genome-wide correlation analyses were performed to identify genes that are significantly correlated with CD47 expression in the human and murine vascular tissue collections described above. These lists were intersected to identify genes that are commonly co-expressed across multiple data sets in both species. The resulting list of genes (Extended Data Fig. 7a) was subjected to a series of bioinformatics analyses including the Database for Annotation, Visualization and Integrated Discovery (DAVID), Kyoto Encyclopedia of Genomes and Genes (KEGG), Gene Ontology (GO), and Protein Analysis Through Evolutionary Relationships (PANTHER) classification. Additionally, genes were mapped to open chromatin regulatory intervals in primary human vascular cells and analysed using the Genomic Regions Enrichment of Annotations Tool (GREAT). Pathways found to be over-expressed were ranked by *P* value. Statistical overrepresentation and enrichment analyses were performed using Bonferroni correction for multiple testing ($P < 0.05$ cutoff).

Ingenuity Pathway Analysis (IPA) was then performed on the resulting intersected CD47 co-expression gene list from human carotid atherosclerosis (BiKE study) and murine aortic atherosclerosis (HMDP study). Briefly, the 63 gene identifiers and expression values were analysed using the Core pathway analysis after removing any duplicates and unmapped identifiers. The resulting networks were then subjected to an Upstream Regulators Analysis using the Ingenuity Knowledge Base after applying a filter to include all genes, RNAs and proteins, while excluding chemicals or drugs. Overlap *P* values were calculated for each upstream regulator using a Fisher's exact test and activation *z*-scores were calculated by comparing observed direction of target genes with inferred literature-derived regulatory direction to identify the most significant upstream regulators for CD47.

Promoter analysis. Using the UCSC browser, the genetic sequence 1 Kb upstream of the CD47 promoter was identified and analysed, as previously described¹⁷. In addition to evaluating for open chromatin status and DNase hypersensitivity sites, potential transcription factor binding sites (TFBS) were predicted using the following online bioinformatics tools: TRANSFAC (BIOBASE), P-Match, TFSearch, Alibaba, PROMO, and MatInspector. High confidence binding sites (85% likelihood cutoff) were accepted for additional analysis. Additionally, the predicted binding sites in the CD47 promoter region were intersected with open chromatin peaks identified from the assay for transposase accessible chromatin followed by sequencing (ATAC-seq) in primary human coronary artery SMCs (HCASMC).

Cell culture. Primary vascular SMCs were collected from the aortas of C57BL/6 mice and propagated in DMEM supplemented with 10% FBS, as previously described^{17,40}. Human coronary artery SMCs (HCASMC, Lonza Catalog CC-2583, passage 3–6) were propagated in SmGM-2 growth media (Lonza) containing 5% FBS. To obtain human macrophages, leukocyte reduction system (LRS) chambers were obtained from the Stanford Blood Center from anonymous donors. Monocytes were purified on an autoMACS Pro Separator (Miltenyi) using whole-blood anti-CD14 microbeads (Miltenyi) and differentiated to macrophages by culture for 7–10 days in IMDM+GlutaMax (Invitrogen) supplemented with 10% AB-human serum (Gemini Bio-Products 100-512) and 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (Invitrogen). RFP⁺ mouse macrophages were generated and evaluated as previously described⁴¹. Briefly, bone-marrow cells were isolated from C57BL/Ka Rosa26 mRFP1 transgenic mice and differentiated in IMDM+GlutaMax supplemented with 10% fetal bovine serum, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin, and 10 ng ml⁻¹ murine M-CSF (PeproTech). Mouse yolk sac endothelial cell line (C166, ATCC, CRL-2581) and mouse macrophage cell line (RAW 264.7, ATCC, TIB71) were grown in DMEM-growth media containing 10% FBS, while mouse T-lymphocyte cell line (EL4, ATCC, TIB-39) were grown in DMEM containing 10% horse serum. Human macrophage cell line (THP1, ATCC, TIB-202) were grown in RPMI-1640 medium containing 10% FBS and 0.05 mM 2-mercaptoethanol. Human embryonic kidney cells (HEK-293, ATCC, CRL-1573) used for luciferase reporter assays were grown in DMEM-growth media containing 10% FBS. No additional cell authentication or mycoplasma contamination testing was performed.

A variety of atherosclerosis-related or pro-apoptotic stimuli were applied to the cells in the experiments described below including: oxidized LDL (oxLDL, 50 µg ml⁻¹, Alfa Aesar), Ang II (100 nM, Sigma-Aldrich), fibroblast growth factor (FGF, 100 ng ml⁻¹, R&D), platelet-derived growth factor (PDGF, 100 ng ml⁻¹, R&D), and lipopolysaccharide (LPS, 1 µg ml⁻¹, Sigma-Aldrich). Additionally, a number of cytokines associated with CD47 expression through the HMDP Luminex array were also tested, including tumour necrosis factor α (TNF-α, 50 ng ml⁻¹, R&D), interleukin 2 (IL-2, 100 ng ml⁻¹, Biolegend), chemokine receptor ligand 1 (CXCL1, 100 ng ml⁻¹, Biolegend), interleukin 4 (IL-4, 50 ng ml⁻¹, Biolegend), and transforming growth factor β (TGF-β, 50 ng ml⁻¹, R&D). Before experimentation, SMCs were serum-starved for 24 h in DMEM and then stimulated with the stimuli listed above for 24 h before analysis. In some experiments, the cells were stimulated with staurosporine (STS, 1 µM, Sigma-Aldrich) for 1 or 4 h to induce apoptosis, after 24 h of TNF-α treatment.

To inhibit TNF-α signalling, a chemical inhibitor (SPD 304, Sigma Aldrich) or a monoclonal antibodies (Infliximab, Janssen) were used. Briefly, TNF-α was pre-incubated with 10 µM of SPD304 or 100 µg ml⁻¹ of infliximab in serum free DMEM for 20 min before cell stimulation. To inhibit NF-κB signalling, the cells were pre-treated with 10 µM BAY 11-7085 (Santa Cruz Biotechnologies) or DMSO, then stimulated with TNF-α.

For MAPK western blotting experiments, mouse aortic SMCs were serum-starved for 48 h, pre-treated with 10 µg ml⁻¹ of CD47 antibodies or IgG for 20 min, then stimulated with thrombospondin-1 (TSP1, 10 µg ml⁻¹, R&D) for 10 or 30 min. For eNOS western blotting experiments, C166 cells were serum-starved for 8 h, pre-treated with 2 µM TSP1 with or without 10 µg ml⁻¹ of CD47 antibody or IgG

for 20 min, then stimulated with acetylcholine (Ach, 10 µM, Sigma-Aldrich) for 15 min.

mRNA isolation and quantitative reverse-transcription PCR. RNA was isolated from cell lysates using the miRNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. RNA was isolated from murine organ samples using the Trizol method (Invitrogen). RNA was quantified with the Nanodrop machine (Agilent Technologies). For quantitation of gene transcription, cDNA was generated with MultiScribe reverse transcriptase (Applied Biosystems), and then amplified on the ABI PRISM 7900HT with commercially available TaqMan primers (Applied Biosystems) and normalized to 18S internal controls, as previously described¹⁷. A list of the primers and probes used in these studies is provided in Extended Data Table 1e.

Protein extraction and western blotting. Total protein was isolated from cultured cell lines and tissue homogenates using 1× cell lysis buffer (Cell Signaling) supplemented with 1× Halt Protease & Phosphatase Single-Use Inhibitor Cocktail (Thermo Scientific), as previously described¹⁷. The protein concentration in each sample was measured using Pierce BCA Protein Assay Kit (Thermo Scientific). Equal amounts of protein were loaded and separated on precast gels (Bio-Rad) and thereafter transferred onto PVDF membranes (Bio-Rad). Following a 1 h incubation in 5% bovine albumin serum solution prepared in 1× TBST, these membranes were probed with commercially available antibodies designed to recognize endogenous P38 (Cell Signaling, 1:1000), phospho-P38 (Cell Signaling, 1:1000), ERK1/2 (Cell Signaling, 1:1000), phospho-ERK1/2 (Cell Signaling, 1:1000), phospho-eNOS (Cell Signaling, 1:1000), CD47 (Novus Biologicals, 1:1000) and GAPD (1:1000; Cell Signaling Technologies) overnight at 4°C. Membranes were rinsed with TBST and incubated with appropriately matched horse-radish peroxidase (HRP)-conjugated anti-mouse (1:5000; Life Technologies) or anti-rabbit (1:5000; Life Technologies) antibodies for 1 h, before protein expression was detected using SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific). Membranes were then scanned with a Licor Odyssey Fc imager for quantitative analysis. In some experiments, membranes loaded with protein were incubated with anti-CD47 antibody that had been preabsorbed with CD47 peptide (R&D systems, 1866-CD, 1:5 dilution) for 16 h, to determine the specificity of the primary antibody.

Apoptosis assays. To evaluate apoptosis, the luminometric Caspase-Glo 3/7 Assay (Promega, G8090) was performed on cultured cells, according to the manufacturer's protocol. Briefly, mouse aortic SMCs were seeded in 96-well plates at the density of 10,000 cells per well, grown at 37°C for 24 h, and then serum-starved for 24 h. Apoptosis was induced with 1 µM STS treatment for 4 h in the presence of 10 µg ml⁻¹ of anti-CD47 antibodies or IgG. Confirmatory assays were performed by flow cytometry, where cells were exposed to 24 or 72 h of vehicle, 50 ng ml⁻¹ of TNF-α, or 50 µg ml⁻¹ of oxLDL, then treated with 1 µM STS for 4 h before being collected in TrypLE. These cells were stained with anti-annexin V antibody labelled with fluorescein isothiocyanate (FITC) and propidium iodide (eBioscience) and analysed by Scanford cell analyser (Stanford Shared facility, Stanford), as previously described⁴². These FACS data was analysed by FlowJo 10.1r5.

Proliferation assays. A modified MTT (3-[4,5-dimethyl-thiazol-2-yl]-2,5-diphenyltetrazolium bromide) assay was performed to analyse SMC proliferation and viability. Mouse aortic SMCs were seeded in 96-well plates at the density of 6,000 cells per well, grown at 37°C overnight, and then serum-starved for 48 h. The cells were stimulated with 10% serum or 10 µg ml⁻¹ TSP1 with or without 10 µg ml⁻¹ of anti-CD47 antibodies or IgG for 24 h and then incubated for 4 h in the presence of 10 µl of MTT AB solution (Millipore). The formazan product was dissolved by addition of 100 µl acidic isopropanol (0.04 N HCl) and absorbance was measured at 570 nm (reference wavelength 630 nm) on an ELISA plate reader by SpectraMax 190 Microplate Reader (Molecular Devices).

Efferocytosis assay. Standard *in vitro* phagocytosis assays were performed as previously described⁴¹. SMCs were labelled with 2.5 µM carboxyfluorescein succinimidyl ester (CFSE) according to the manufacturer's protocol (Invitrogen). 100,000 SMCs were plated per well in a 96-well plate (Corning) and pre-incubated with antibody (IgG1 isotype control (MOPC-21) or M1AP410 (anti-CD47)) for 30 min at 37°C. An unrelated anti-CD8 antibody was also tested as a negative control. After 30 min, 50,000 macrophages were added to each well and co-incubated for 2 h in serum-free medium, then analysed using an LSRFortessa cell analyser with high throughput sampler (BD Biosciences). RFP⁺ mouse macrophages were identified by intrinsic fluorescence. Dead cells were excluded from the analysis by staining with DAPI (Sigma). Phagocytosis was evaluated as the percentage of GFP⁺ macrophages using FlowJo X 10.0.7r2 (Tree Star) and was normalized to the maximal response by each independent donor against each cell line. In addition to measuring basal efferocytosis rates, experiments were repeated with target cells that had been pretreated with a variety of compounds (alone or in combination) including STS, oxLDL, TNF-α, TSP1, and infliximab. Confirmatory assays were

performed with RAW macrophages where target cells were labelled with 1 μ M of CellTracker Deep Red dye (Life technologies) and phagocytes were labelled with 1.25 μ M of CellTracker Orange CMRA (Life technologies) at 37 °C for 30 min. As above, cells in these assays were treated with 50 μ g ml⁻¹ of oxLDL, 50 ng ml⁻¹ of TNF- α , or 100 μ g ml⁻¹ of infliximab antibody for 24 h before co-culture in serum-free medium with 10 μ g ml⁻¹ of antibody of IgG1 isotype control (MOPC-21) or anti-CD47 (MIAP410) for 2 h at 37 °C. Double-positive cells were quantified using the Scanford cell analyser (Stanford Shared facility) and analysed by FlowJo 10.1r5, as previously described¹⁷. Statistical significance was determined by one-way or two-way ANOVA with Bonferroni's correction using Prism 5 (Graphpad).

Flow cytometry. To measure the cell surface expression of CD47, cells were exposed to vehicle or 50 ng ml⁻¹ of TNF- α for 24 h. In some experiments, cells were treated with 1 μ M STS during the last 1 h or 4 h of the incubation, before analysis. The cells were collected in TrypLE (Life Sciences) and stained with anti-CD47 antibodies (AbD Serotec MCA2514GA, clone1/1A4, 1:400) or IgG (eBioscience), followed by Alexa Fluor 488 goat anti-mouse (Life Technologies, 1:400), and then FACS-sorted within 1 h (BD FACSCaliber, 530 nm fluorescence [FL1] and >575 nm [FL3]). Analysis was performed with FloJo 7.6.3.

Immunocytochemistry. Primary mouse aortic SMCs were seeded at approximately 60% confluence in glass-bottom culture dishes (MatTek Corporation). Following treatment with TNF- α and/or STS, as described above, cells were rinsed with PBS and fixed with freshly prepared 4% PFA (Fisher Scientific). Once permeabilized with 0.1% Triton X-100 (Sigma), cells were incubated with blocking buffer (3% BSA, Cell Signaling Technology) for 1 h and then incubated overnight at 4 °C with a CD47 antibody (1:80; R&D Systems, Cat# AF1866-SP). After rinsing with PBS, cells were incubated in the dark with a donkey anti-goat Alexa Fluor 594 conjugate secondary antibody (1:1000; Life Technologies) for 1 h then briefly incubated with DAPI. CD47 and DAPI localization was captured at 20 \times magnification using a Leica DMI3000 B microscope capable of taking fluorescent images. Studies using primary human coronary artery SMCs were also performed as above, but used 5% goat serum blocking buffer (ThermoFisher Scientific), and the following primary and secondary antibodies: CD47 antibody (Novus Biologicals, 1:50); HMGB1 (abcam, ab18256 1:100); Alexa Fluor 594 goat anti-mouse (Life Technologies A11005, 1:300); and Alexa Fluor 488 goat anti-rabbit (Life Technologies, A11034). In some studies, exogenous CD47 peptide was preincubated with the cells before the CD47 antibody was applied, as described above.

Luciferase reporter assay. CD47 LightSwitch Promoter Reporter GoClones (RenSP, S710450), empty promoter vectors (S790005) and Cypridina TK Control constructs (pTK-Cluc, SN0322S) were obtained from SwitchGear Genomics and transfected into HEK cells using lipofectamine 2000 (Invitrogen). For overexpression assays, expression plasmids for *Nfkb1* (p50), *Rela* (p65), *Nfkb2* (p52), and *c-Rel* were obtained from Addgene (#21965, #21966, #23289, #27256, respectively). Empty vector pCMV4 was generated by HindIII digestion of #21966, and pcDNA3.1 was obtained from Life Technologies. 50 ng of plasmid was co-transfected with 45 ng of the RenSP reporter and 5 ng of the pTK-Cluc reporter construct. Media was changed to fresh DMEM and 50 ng ml⁻¹ of TNF- α was added 2, 8, 24, and 36 h before collection. The cell lysate and supernatant were collected 48 h after transfection and dual luciferase activity was measured with the LightSwitch Dual Assay System using a SpectraMax L luminometer (Molecular Devices), according to the manufacturer's instructions. Relative luciferase activity (*Renilla/Cypridina* luciferase ratio) was quantified as the percentage change relative to the basal values obtained from control-transfected cells not exposed to TNF- α treatment.

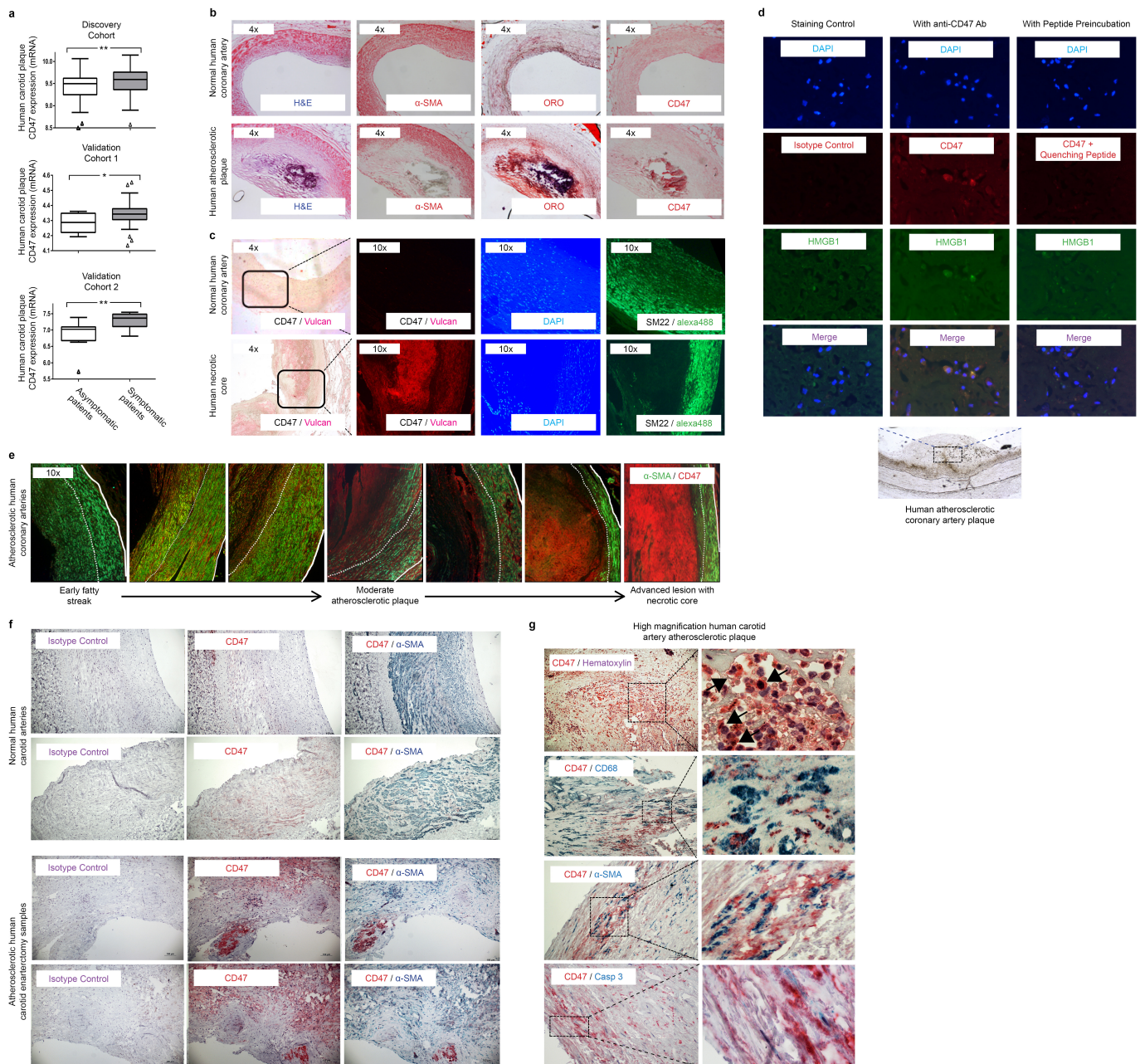
Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) was performed according to the Millipore Magna-ChIP protocol with slight modifications. HCASMC were cultured in normal growth media until approximately 75% confluent. Cells were fixed in 1% formaldehyde for 10 min to cross-link chromatin, followed by quenching with glycine for 5 min at room temperature. 2 \times 10⁷ cells per condition were collected, and nuclear lysates were prepared according to the manufacturer's protocol. Cross-linked chromatin nuclear extracts were sheared into approximately 500 bp fragments using a Bioruptor (Diagenode) for 3 cycles of 3 min (30 s on, 30 s off). Sheared chromatin was clarified via centrifugation at 4 °C for 10 min. 1 \times 10⁶ nuclei per condition was incubated with 2 μ g rabbit IgG or anti-NF- κ B p105/p50 antibody (Abcam, Ab7971) plus protein A/G magnetic beads overnight at 4 °C on a rotating platform to capture the protein-DNA complexes. Complexes were washed in low salt, high salt, LiCl, and TE buffers and then eluted with a ChIP Elution Buffer containing Proteinase K. Free DNA was subsequently purified using spin columns. Total enrichment was measured using primers designed based on the sequence of the top *Nfkb* binding site within the CD47 promoter: forward (-804 to -781):

5'-ATAGGGAAGAGCAGAGCGAGTAGA-3' and reverse (+627 to +609): 5'-GCGTGGACCAGGACACCTA-3'), or a negative control region using the following primers: forward: 5'-CCGGAAGCACTTCTCTAGTA-3' and reverse: 5'-AAGAGAGAGCGGAAGTGACG-3'. Quantitative real-time PCR (ViiA 7, Life Technologies) was performed using SYBR Green (Applied Biosystems) assays and fold enrichment was calculated by measuring the $\Delta\Delta C_t - \Delta\Delta C_t$ IgG. Melting curve analysis was also performed for each ChIP primer. Data are presented as the percentage of input DNA and as fold enrichment of chromatin precipitated with the NF- κ B antibodies relative to the control IgG. In some experiments, cells were treated with TNF- α for 90 min and 24 h before isolation of nuclear lysates.

Statistical analysis. Aside from the microarray data (described above), all experimental data are presented as mean \pm s.e.m. Data were subjected to the Kolmogorov-Smirnov test to determine distribution. Groups were compared using the Mann-Whitney *U* test for non-parametric data or the two-tailed Students *t*-test for parametric data. When comparing multiple groups, data were analysed by analysis of variance with one way ANOVA followed by Tukey's or Dunnett's post-hoc test. For multiple testing of parametric data, a value of *P* < 0.05 was considered statistically significant. *In vitro* experiments were replicated at least in triplicate and all analyses were performed in a blinded fashion by two separate investigators, unless otherwise specified. In the *in vivo* intervention studies, comparison was made between mice treated with anti-CD47 antibodies and IgG. In the *in vivo* synergy studies, ANOVA was performed as above with multiple comparison and linear trend post-testing across all four groups. In the *in vitro* studies, comparison was made between the intervention (anti-CD47 antibodies) and control (IgG) arms. 'Vehicle control' was only used in experiments where an additional treatment (for example, TNF, oxLDL, infliximab) was studied. In the TaqMan-based CD47 expression experiments, changes were demonstrated as mRNA fold-change compared to the baseline condition (set as '1'). In the microarray-based experiments, relative expression differences were displayed across conditions (for example, atherosclerosis versus no atherosclerosis). In the *in vitro* phagocytosis assays, efferocytosis rates are displayed as percent of maximum, as previously described⁴¹. Statistical analysis was performed with GraphPad Prism 5. Aside from the human plaque microarray studies (displayed as Tukey box plots) and the correlation plots (displayed as the 95% confidence band of the best fit line), all error bars display the standard error of the mean.

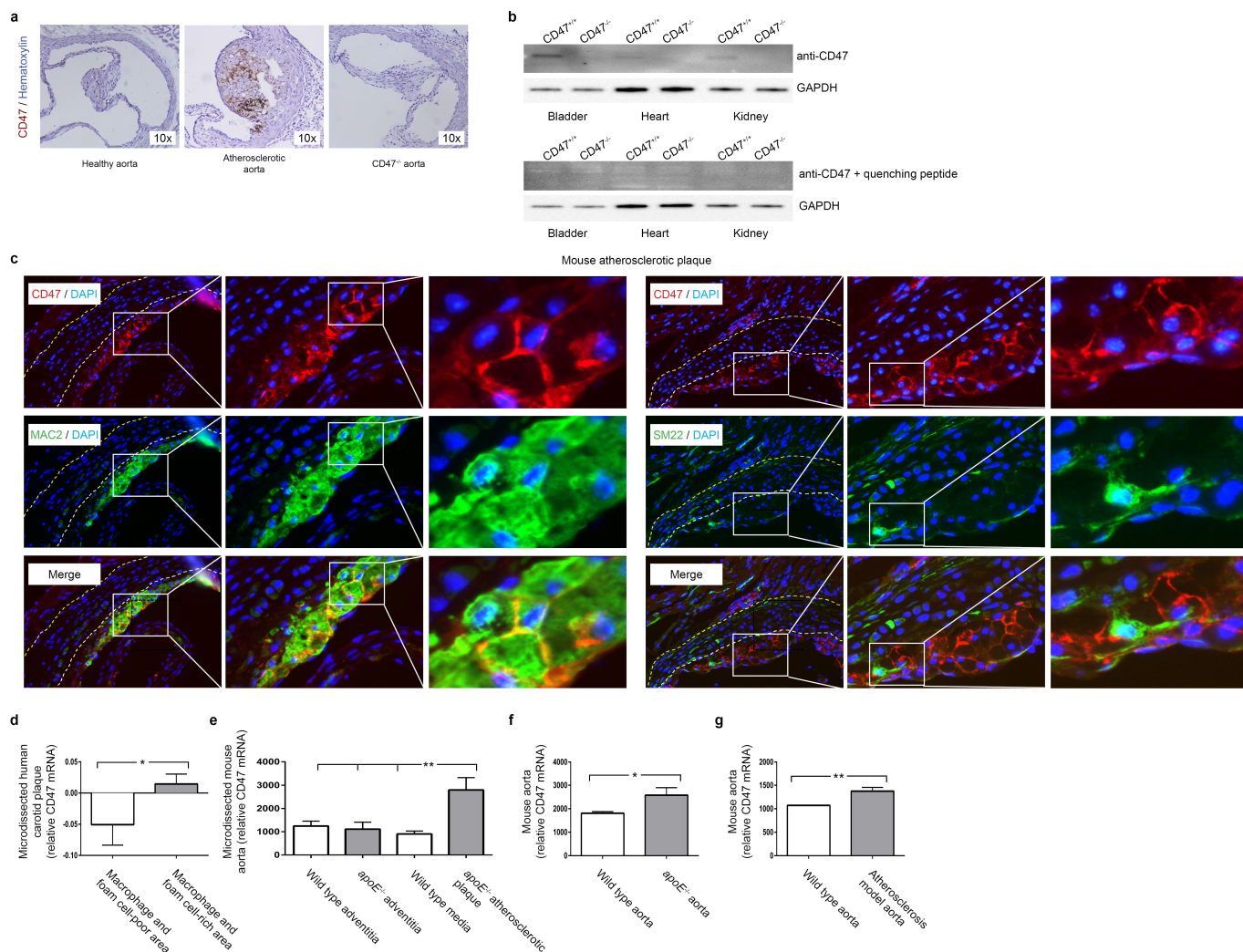
Study approval. All animal studies were approved by the Stanford University Administrative Panel on Laboratory Animal Care (protocol 27279) and conform to the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH Publication No. 85-23, revised 1996). All human studies were performed with written informed consent and with the approval of the Ethical Committee of Northern Stockholm (BiKE).

- Saksi, J. *et al.* Gene expression differences between stroke-associated and asymptomatic carotid plaques. *J. Mol. Med. (Berl.)* **89**, 1015-1026 (2011).
- King, J. Y. *et al.* Pathway analysis of coronary atherosclerosis. *Physiol. Genomics* **23**, 103-118 (2005).
- Ashley, E. A. *et al.* Network analysis of human in-stent restenosis. *Circulation* **114**, 2644-2654 (2006).
- Perisic, L. *et al.* Gene expression signatures, pathways and networks in carotid atherosclerosis. *J. Intern. Med.* **279**, 293-308 (2016).
- Chen, Y. C. *et al.* A novel mouse model of atherosclerotic plaque instability for drug testing and mechanistic/therapeutic discoveries using gene and microRNA expression profiling. *Circ. Res.* **113**, 252-265 (2013).
- Jun, J. *et al.* Effect of intermittent hypoxia on atherosclerosis in apolipoprotein E-deficient mice. *Atherosclerosis* **209**, 381-386 (2010).
- Alexander, M. R. *et al.* Genetic inactivation of IL-1 signaling enhances atherosclerotic plaque instability and reduces outward vessel remodeling in advanced atherosclerosis in mice. *J. Clin. Invest.* **122**, 70-79 (2012).
- Yang, H. *et al.* A customized and versatile high-density genotyping array for the mouse. *Nat. Methods* **6**, 663-666 (2009).
- Ghazalpour, A. *et al.* Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet.* **7**, e1001393 (2011).
- Rong, J. X., Shapiro, M., Trogen, E. & Fisher, E. A. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *Proc. Natl Acad. Sci. USA* **100**, 13531-13536 (2003).
- Weiskopf, K. *et al.* Engineered SIRP α variants as immunotherapeutic adjuvants to anticancer antibodies. *Science* **341**, 88-91 (2013).
- Leeper, N. J. *et al.* Loss of CDKN2B promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arterioscler. Thromb. Vasc. Biol.* **33**, e1-e10 (2013).
- Hägg, S. *et al.* Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm Atherosclerosis Gene Expression (STAGE) study. *PLoS Genet.* **5**, e1000754 (2009).



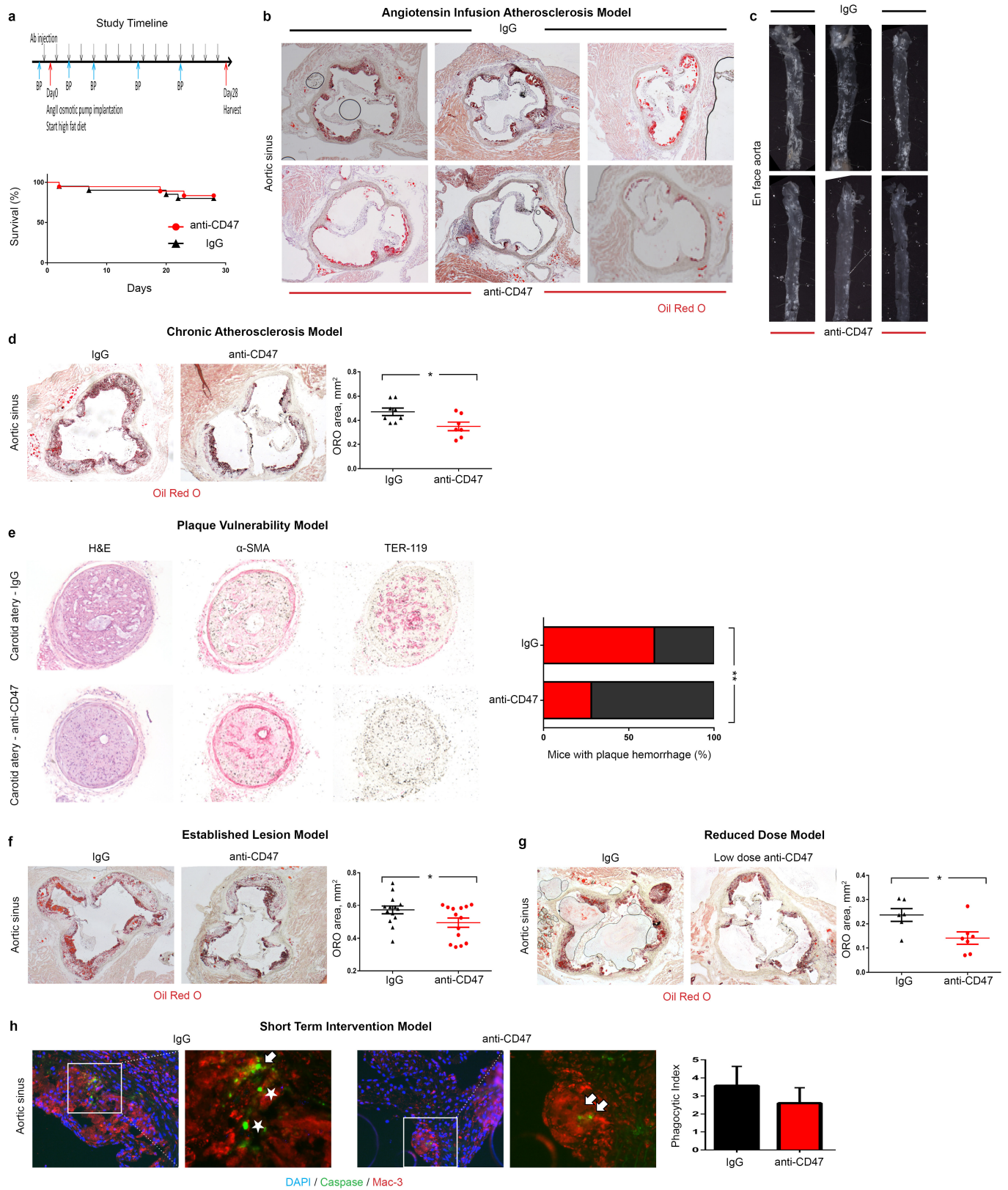
Extended Data Figure 1 | CD47 expression correlates with risk for clinical cardiovascular events and is progressively upregulated in the necrotic core of human blood vessels during atherogenesis. **a**, cDNA microarray expression profiling in the BiKE carotid endarterectomy biobank reveals that the relative expression of CD47 is increased in vascular homogenates taken from subjects with symptomatic disease (stroke or transient ischaemic attack, $n = 85$) compared to those with stable, asymptomatic lesions ($n = 40$). Similar findings were observed in the non-overlapping discovery and validation cohorts from BiKE ($n = 55$), and a second validation cohort from the Helsinki Carotid Endarterectomy Study (HeCES, $n = 21$). Data presented as Tukey box plots. **b**, Immunohistochemical staining reveals that CD47 co-localizes with lipidated plaque within human coronary lesions, as measured by Oil-Red-O (ORO) staining. **c**, Immunofluorescence staining of coronary samples confirms that CD47 is upregulated within the necrotic core. **d**, High magnification (40x) imaging of atherosclerotic coronary plaque confirms that CD47 expression is present on the surface of nucleated cells

undergoing cell death, as indicated by HMGB1 staining. Specificity of the anti-CD47 antibody is confirmed in assays where the signal was quenched by preincubating the sections with recombinant CD47 peptide before primary antibody exposure. **e**, Additional representative coronary artery segments spanning the spectrum of progressive coronary artery disease (non-atherosclerotic coronary, early 'fatty streak, inwardly remodelled plaque, and advanced ulcerated lesion with necrotic core) confirm that CD47 is progressively upregulated during the development of coronary artery disease. The tunica media is indicated by dotted lines. **f**, Additional staining in human carotid artery sections confirms that CD47 expression is upregulated in atherosclerosis relative to healthy tissue, and appears most pronounced within the necrotic core. **g**, High magnification (100x) imaging confirms that the CD47 expression is specific to lesional cells, including SMCs (α-SMA), macrophages (CD68) and cells undergoing programmed cell death (Casp3). Comparisons made by two-tailed t -tests. $^{**}P < 0.01$, $^{*}P < 0.05$. Original magnification, $\times 40$ (**d**), $\times 4$ (**f**, **g**).



Extended Data Figure 2 | CD47 expression is increased in mouse models of atherosclerosis. **a**, Mice injected with biotin-labelled anti-CD47 antibodies reveal that these antibodies accumulate in the vasculature of atherosclerotic mice (middle), relative to non-atherosclerotic control mice (left). No staining is detected in *Cd47*^{-/-} mice (right), indicating specificity of the antibody. **b**, Western blotting of tissue homogenates obtained from wild-type and *Cd47*^{-/-} mice (with and without quenching CD47 peptide) further confirms the specificity of the antibody. For gel source data, see Supplementary Fig. 1. **c**, High resolution immunofluorescence staining of murine atherosclerotic plaques indicate that CD47 is specifically expressed on the surface of lesional

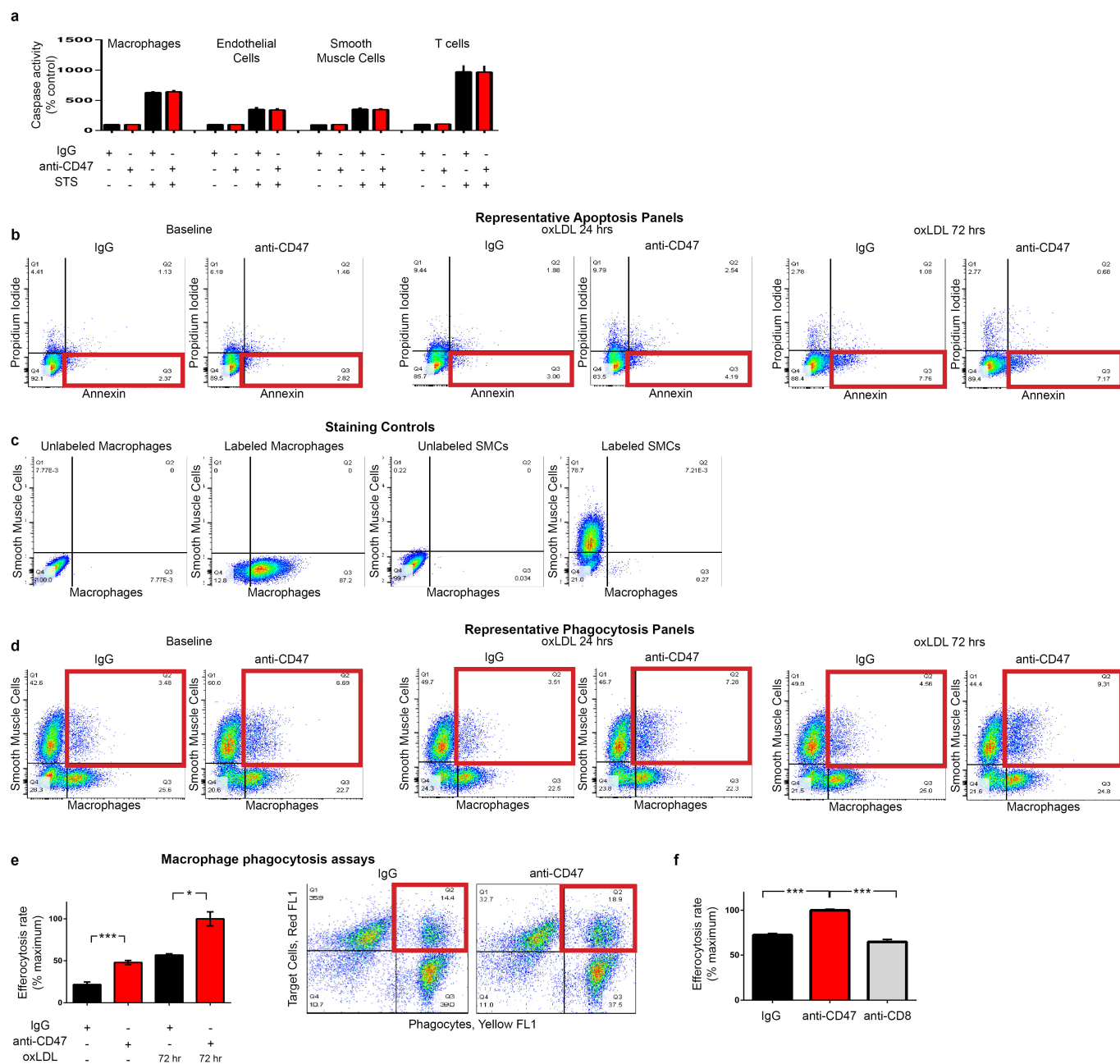
cells, rather than extracellular debris. Original magnification, $\times 40$. **d**, Publically available microarray data from laser capture microdissected (LCM) vascular tissue reveals that CD47 expression is increased within the macrophage and foam-cell-rich area of human plaque, relative to macrophage and foam-cell-poor areas (GSE23303). **e**, Similar results were observed in LCM tissue from mouse atherosclerotic plaque tissue, relative to non-atherosclerotic medial and adventitial tissue (GSE21419). **f**, **g**, Additional results from the Gene Expression Omnibus (GEO) database reveal that aortic CD47 expression is upregulated in murine models of atherosclerosis, as observed in the current study (GSE2372 and GSE19286). ** $P < 0.03$, * $P < 0.05$.



Extended Data Figure 3 | See next page for caption.

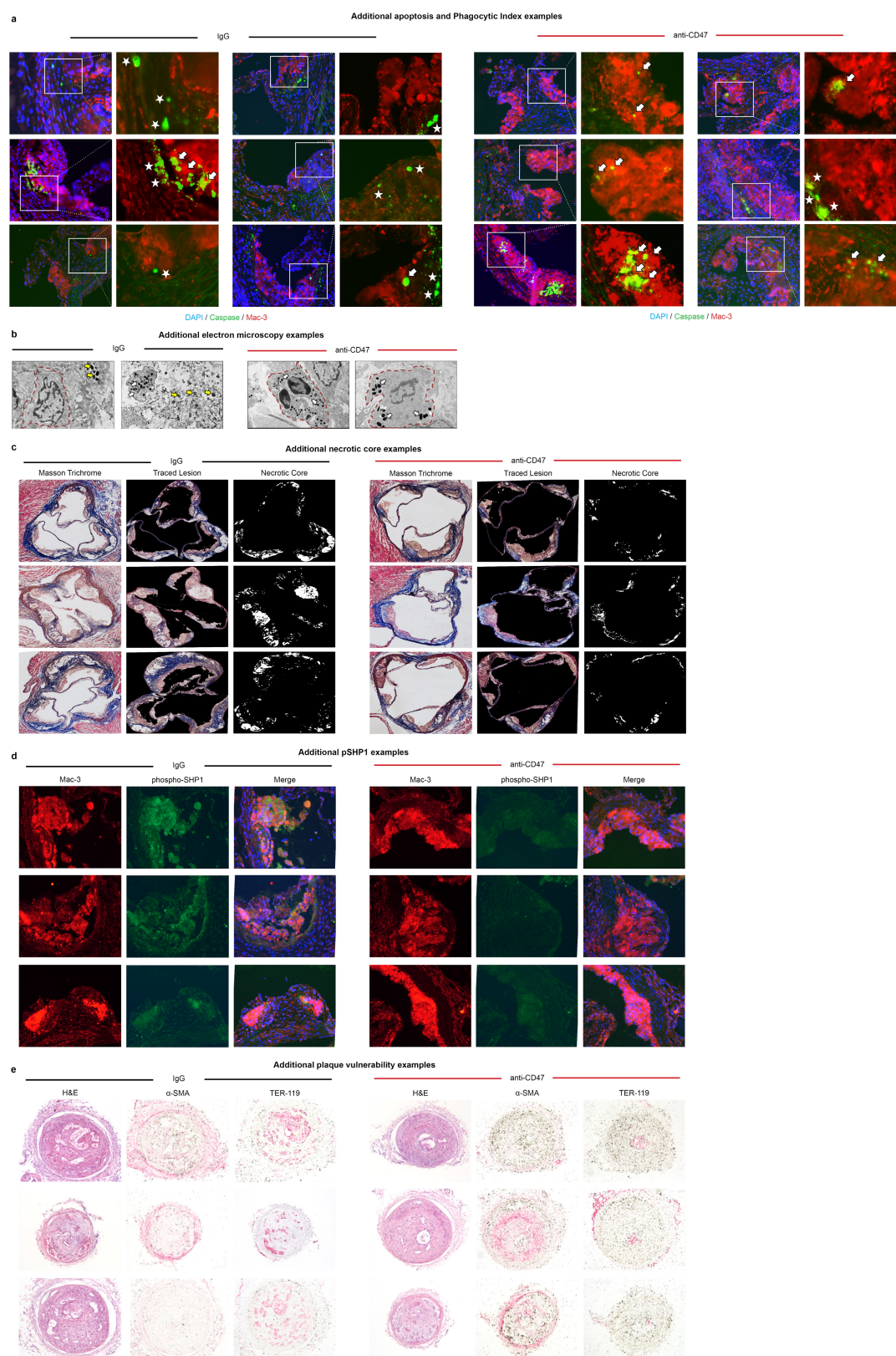
Extended Data Figure 3 | Anti-CD47 antibody reduces atherosclerotic burden in several orthogonal *in vivo* models. **a**, Study timeline detailing osmotic minipump implantation and high-fat feeding to induce atherosclerosis in the *apoE*^{-/-} and ‘angiotensin infusion’ model used herein. Kaplan–Meier curves indicate no change in mortality with anti-CD47 treatment during 28 days of follow up. **b**, **c**, Additional representative examples confirm that anti-CD47 antibody reduces atherosclerosis content in the aortic sinus (**b**) and reduces the percentage of the *en face* aorta covered by atherosclerotic plaque (**c**). **d–g**, Several additional atherosclerosis models were also used in this study to confirm the beneficial effects of anti-CD47 antibody therapy, and to model additional aspects of human cardiovascular disease, including a ‘chronic atherosclerosis’ model, where antibody therapy was given for 12 weeks (with no angiotensin infusion) (**d**); a ‘plaque vulnerability’ model, where

the impact of antibody therapy on plaque rupture and intraplaque haemorrhage was quantified (**e**); an ‘established disease’ model, where therapy was given for 7 weeks after mice had already developed advanced plaques of equivalent size (**f**); and a ‘reduced dose’ model, where the dose of anti-CD47 antibody was reduced by 75%, relative to the preceding studies (**g**). **h**, Additionally, a ‘short term’ study was performed where mice with established lesions of equivalent size and identical apoptosis rates were pulsed for only 5 days with anti-CD47 antibodies before collection, to quantify the effect of therapy on efferocytosis rates, independent of lesion size (phagocytic index indicated by the ratio of ‘free’ (white stars) to ‘associated’ (white arrows) apoptotic bodies). Additional methodological details are provided in the Methods. Comparisons made by two-tailed *t*-tests. ***P* < 0.03, **P* < 0.05. Error bars represent the s.e.m. Original magnification, ×4 (**b**, **d–f**), ×2 (**c**), ×10 (**h**).



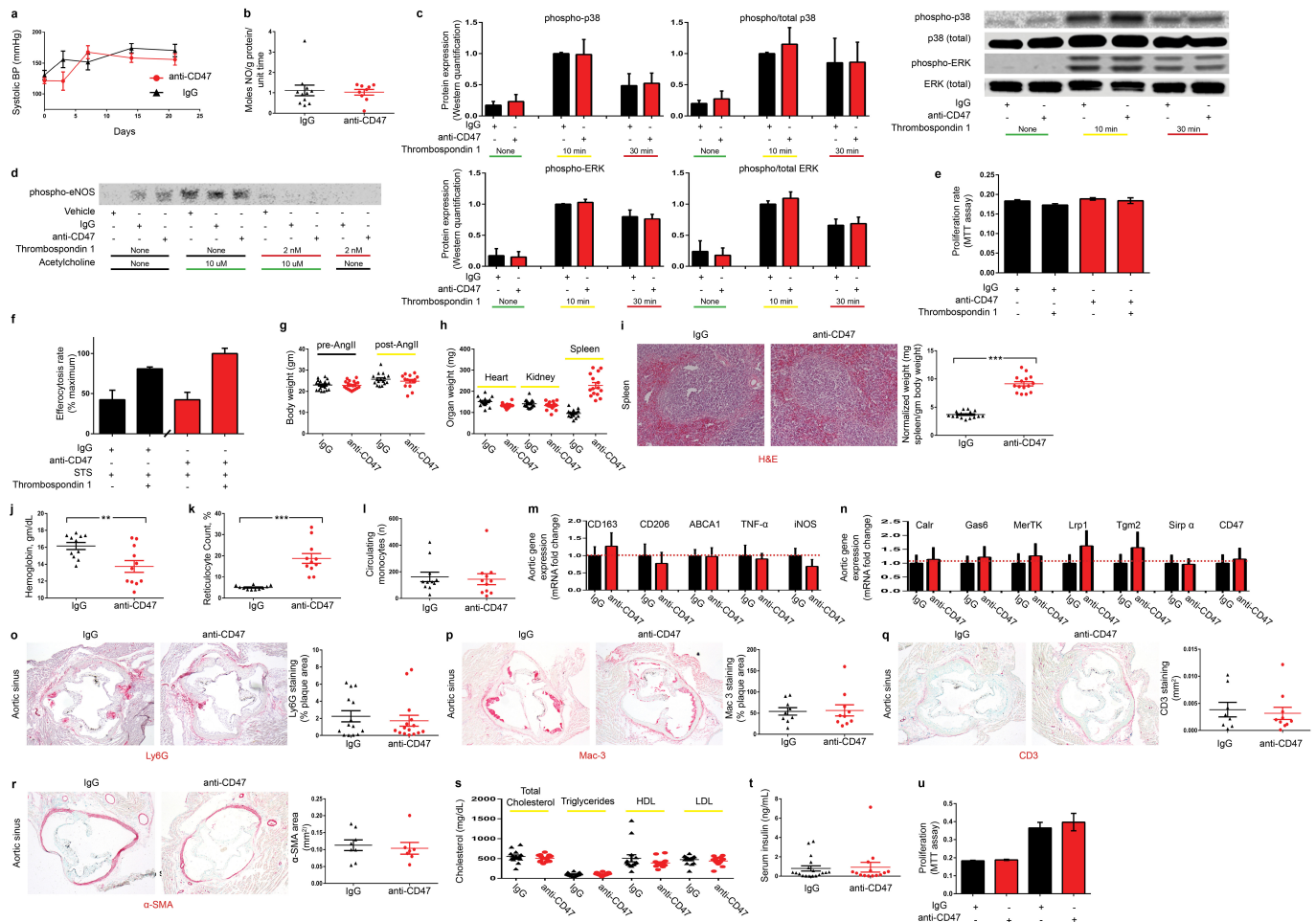
Extended Data Figure 4 | Anti-CD47 antibody promotes the phagocytosis of diseased SMCs and macrophages, without altering apoptosis. **a**, *In vitro* caspase activity assays reveal that anti-CD47 antibody does not alter rates of programmed cell death in any vascular cell type. **b**, Flow cytometry assays confirm that anti-CD47 antibody has no effect on apoptosis at baseline, or in vascular SMCs exposed to 24 or 72 h of oxLDL. **c**, Staining controls for the *in vitro* phagocytosis assays. **d**, Representative FACS plots for the *in vitro* efferocytosis conditions

displayed in Fig. 2e. The right upper quadrant (highlighted in red) includes double-positive cells that are taken to represent a macrophage that has ingested a target cell. **e**, *In vitro* efferocytosis assays using lipid-loaded macrophages as the target cell confirm that anti-CD47 antibody also stimulates the clearance of this vascular cell type, similar to the findings observed with SMCs. **f**, Additional *in vitro* efferocytosis assays confirm that anti-CD47 antibody stimulates phagocytosis of vascular cells in a specific manner. Error bars represent the s.e.m.



Extended Data Figure 5 | Additional examples confirm the pro-ferrocytic properties of anti-CD47 antibody *in vivo*. **a**, Additional representative images detail that mice treated with anti-CD47 antibodies have a lower overall burden of apoptotic debris (caspase in green), as well as fewer examples of 'free' apoptotic bodies (white stars). Those apoptotic bodies that are present in these lesions are more often found in close proximity to macrophages (Mac-3 in red) and are considered 'associated' with a phagocyte if physically co-localized (white arrows). **b**, Additional electron microscopy examples provide further qualitative evidence

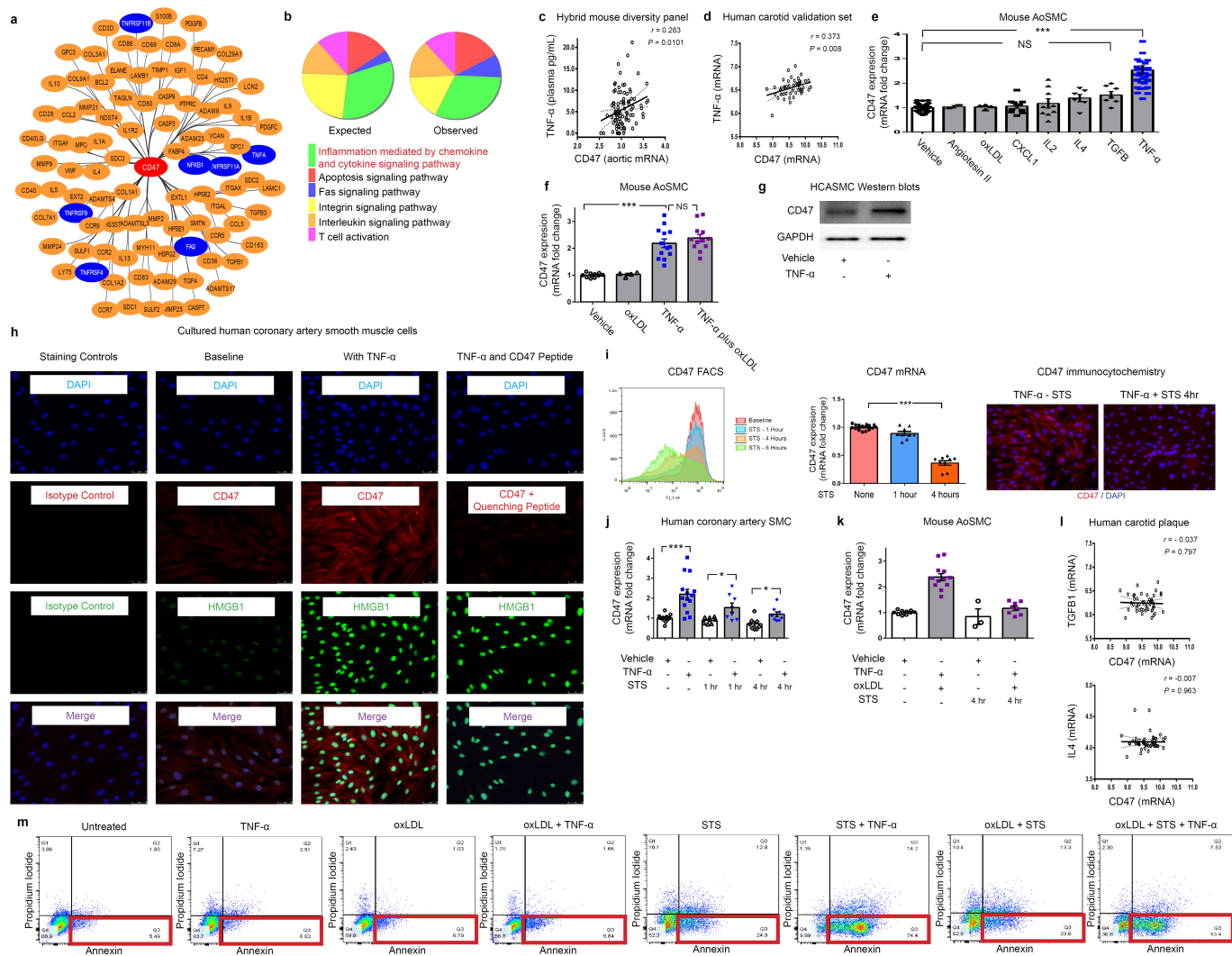
that phagocytes present in the lesions of mice treated with anti-CD47 antibodies are more likely to have ingested several apoptotic bodies (white arrows) compared to lesions from IgG treated mice which are more likely to have a high burden of 'free' apoptotic bodies (yellow arrows). **c–e**, Additional representative examples of the necrotic core analysis (**c**), the phospho-SHP1 staining (**d**), and the plaque haemorrhage analysis (**e**) are shown, as described in the Methods. Original magnification, $\times 10$ (**a**, **d**), $\times 4$ (**c**, **e**).



Extended Data Figure 6 | Full-dose anti-CD47 antibody therapy induces anaemia, but does not appear to alter NO elaboration, TSP1-dependent signalling, or other processes relevant to vascular biology.

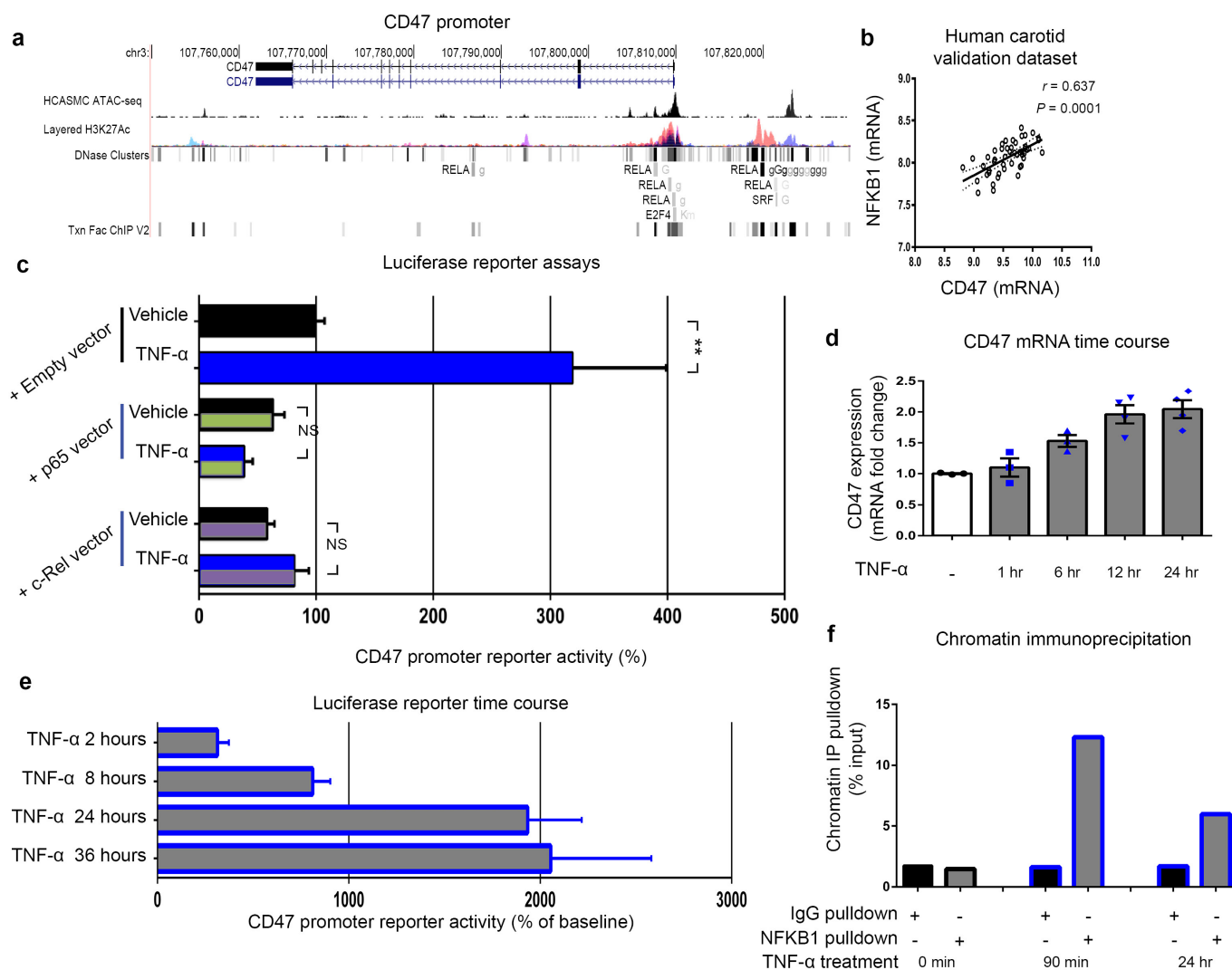
a, No significant change in blood pressure is observed between mice treated with IgG or anti-CD47 antibodies, arguing against a systemic difference in nitric oxide (NO) production due to antibody therapy. **b**, Direct measurement of pulmonary NO release via the Griess reaction indicates that anti-CD47 antibody does not increase NO elaboration *in vivo*. **c**, Western blot analysis of cultured murine vascular cells reveals that anti-CD47 antibody has no effect on the expected induction of p38 and ERK phosphorylation secondary to TSP1 treatment. **d**, Similarly, anti-CD47 antibody has no effect on TSP1-dependent inhibition of eNOS phosphorylation, nor acetylcholine-dependent induction of eNOS phosphorylation. **e**, MTT assays show that anti-CD47 antibody does not affect cellular proliferation rates in the presence of TSP1. **f**, *In vitro* efferocytosis assays show that the expected basal increase in phagocytosis observed after apoptotic cells are exposed to TSP1 (black bars) is not altered in the presence of anti-CD47 antibodies (red bars). **g**, Compared to mice receiving control IgG, mice receiving anti-CD47 antibody treatment have similar body weights at baseline and at time of killing. **h**, No difference is observed for the weight of any organ between groups, with the exception of splenomegaly observed in the anti-CD47-treated animals. **i**, Histological analysis of the explanted splenic tissue reveals an increase in the red pulp of anti-CD47 treated mice without any change in fibrosis or white pulp content, suggestive of increased erythrophagocytosis

in this reticuloendothelial organ. **j–l**, Dot plots detail the haemoglobin count (**j**), reticulocyte count (**k**) and circulating monocyte count (**l**) for each animal in the acute 4-week angiotensin-infusion atherosclerosis model. Note that this anaemia appears to be self-limited, and no anaemia was observed in the chronic atherosclerosis model or the reduced dose model ($P = 0.54$ and 0.57 , respectively). **m**, mRNA analysis of aortic tissue reveals that anti-CD47 antibody has no significant effect on the expression of macrophage-polarization factors *in vivo*. **n**, Anti-CD47 antibody also has no effect on the aortic expression of any other candidate efferocytosis genes. **o–r**, Additional quantitative analyses reveals that anti-CD47 antibody has no effect on *in vivo* neutrophil content (as assessed by Ly6G⁺ area normalized to lesion size) (**o**); macrophage content (as assessed by Mac-3⁺ area normalized to lesion size) (**p**); T-cell content (as assessed by CD3⁺ area across the lesion and adventitia) (**q**); or smooth muscle cell content (as quantified by α-SMA⁺ area in the aortic sinus from the external elastic lamina to the lumen) (**r**). **s, t**, Anti-CD47 antibody also had no effect on lipid level (**s**) or serum insulin (**t**). **u**, MTT assays reveal that anti-CD47 antibody has no effect on the proliferation of primary aortic SMCs obtained from *apoE*^{−/−} mice either at baseline (left) or in the presence of 10% serum (right). Comparisons made by two-tailed *t*-tests, unless otherwise specified. *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$. Error bars represent the s.e.m. For gel source data, see Supplementary Fig. 1; for detailed serological data, see Extended Data Table 1. Original magnification, $\times 4$.



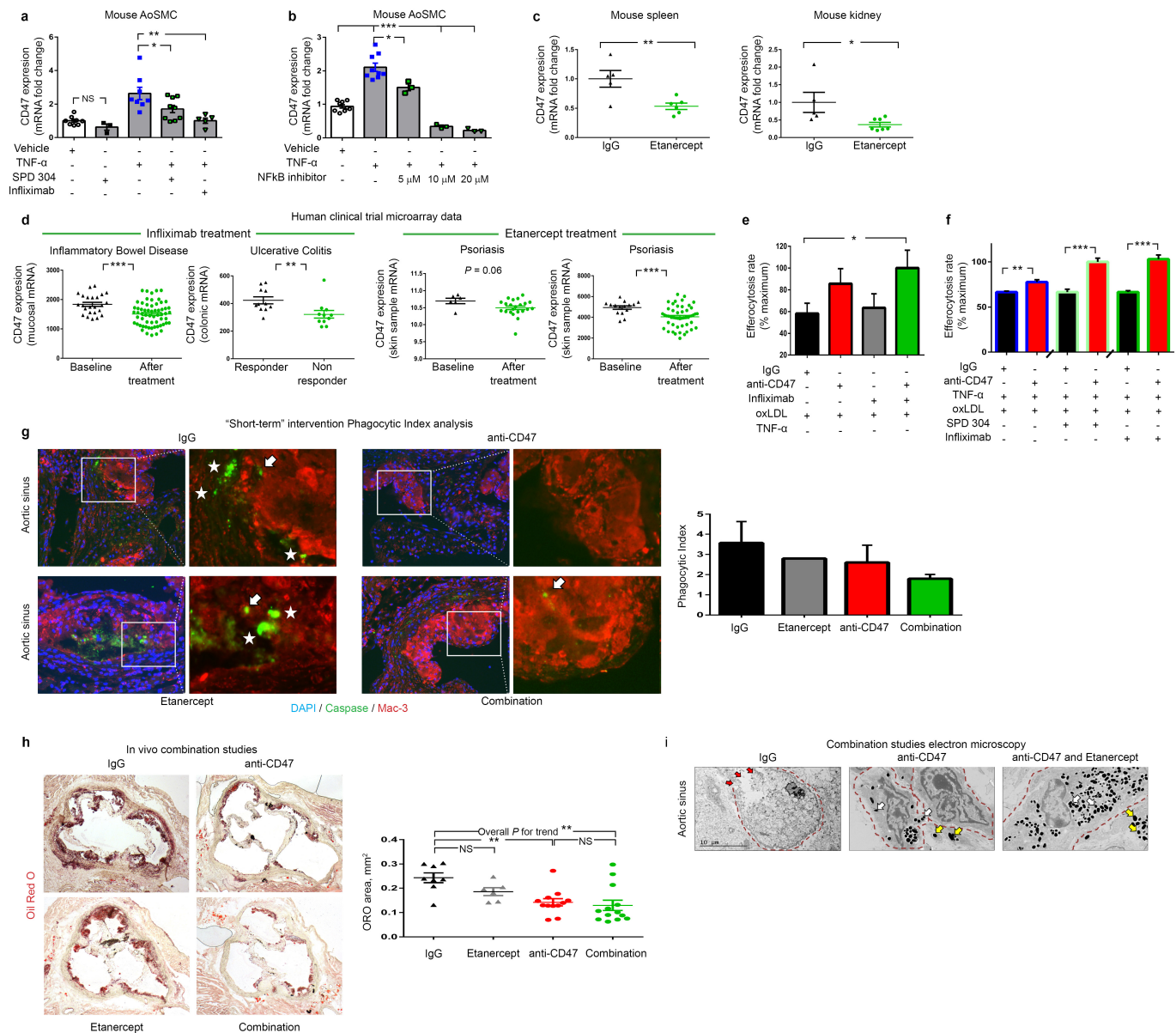
Extended Data Figure 7 | Additional bioinformatic and experimental analyses further implicate a central role for TNF-α in vascular CD47 signalling. **a**, Cytoscape network visualization of the genes which are significantly correlated with CD47 expression in both human and murine atherosclerotic plaque reveals a high number of TNF-α-related factors (indicated in blue), including ligands, receptors, and downstream signalling factors. **b**, PANTHER pathway analysis of those genes which were significantly associated with CD47 expression in mouse and human vascular tissue and have been previously associated with atherosclerosis through the STAGE study⁴³, identifies 'inflammation mediated by chemokine and cytokine signaling pathway' as the most over-abundant pathway associated with CD47 expression in vascular tissue. **c**, Using the Hybrid Mouse Diversity Panel (HMDP), which correlates aortic gene expression with Luminex cytokine array data of plasma samples from over 100 inbred strains of mice, we found that vascular CD47 expression is positively correlated with three inflammatory cytokines *in vivo*, including TNF-α, IL-2 and CXCL1. Correlation data shown for CD47 and TNF-α. **d**, Co-expression studies confirm that TNF-α and CD47 expression are positively correlated in human carotid endarterectomy samples from the BiKE validation study. The Pearson correlation coefficient was determined assuming a Gaussian distribution and *P* values were determined using a two-tailed test. **e**, Experiments with primarily cultured mouse aortic SMCs indicate that TNF-α reproducibly induces *Cd47* mRNA upregulation, whereas a number of other classical pro-atherosclerotic stimuli have no significant effect. Notably, CXCL1, IL4, TGF-β and IL-2 fail to induce CD47 expression *in vitro*, as assessed

by ANOVA. **f**, Additional studies suggest that the effect of TNF-α on CD47 expression persists in the presence of oxLDL, as occurs in the atherosclerotic plaque. **g**, Western blotting confirms that TNF-α induces CD47 expression in vascular cells at the protein level. For gel source data, see Supplementary Fig. 1. **h**, Immunocytochemistry studies of HCASMCs confirm that CD47 expression is induced on the cell surface of TNF-α treated cells. TNF-α effect is assessed by co-staining for HMGB1, and antibody specificity is confirmed with isotype control and recombinant CD47 peptide quenching assays. **i**, Multiple assays (including FACS, TaqMan and immunocytochemistry studies) reveal that CD47 expression is downregulated on vascular SMCs during programmed cell death, as has previously been observed with inflammatory cells. **j**, Confirmatory assays in cultured human coronary artery SMCs reveal that TNF-α induces changes similar to those observed in murine cells (Fig. 3d), including an induction of CD47 under physiological conditions and a blunting of its expected downregulation during apoptosis. Original magnification, $\times 20$. **k**, The capacity of TNF-α to impair CD47 downregulation during programmed cell death is also observed in mouse SMCs simultaneously exposed to pro-apoptotic stimuli and oxLDL. **l**, No correlation between CD47 and other candidate cytokines was observed in the BiKE biobank, further supporting a specific relationship between CD47 and TNF-α. **m**, Representative FACS-based apoptosis panels from cells exposed to the conditions used in Fig. 3g confirm that TNF-α suppresses efferocytosis (Fig. 3g) despite increasing programmed cell death. Comparisons made by two-tailed *t*-tests, unless otherwise specified. *****P* < 0.001, **P* < 0.05. Error bars represent the s.e.m.



Extended Data Figure 8 | The *CD47* promoter contains predicted binding sites for the TNF- α -related transcription factor NF- κ B1. **a**, UCSC genome browser screenshot showing overlay of human *CD47* transcript with ENCODE transcription factor binding sites (including RELA, E2F4, and SRF), along with the active H3K27ac histone modification ChIP-seq track, and a custom track for chromatin accessibility in HCASMCs using the assay for transposase accessible chromatin followed by sequencing (ATAC-seq). These chromatin, DNase hypersensitivity sites, and published ChIP-seq data suggest that members of the NF- κ B family of transcription factors could regulate *CD47* expression in vascular tissue. **b**, Additional co-expression studies in the BiKE validation study confirm that *NFKB1* and *CD47* expression are positively correlated in human carotid endarterectomy samples. The

Pearson correlation coefficient was determined assuming a Gaussian distribution and *P* values were determined using a two-tailed test. **c**, Additional luciferase promoter reporter assays reveal that induction of CD47 expression requires the presence of NF- κ B1 and cannot be induced by other NF- κ B co-factors such as RELA or c-REL. **d, e**, Time-course studies confirm that CD47 expression is induced by TNF- α within 24 h, suggesting a direct transcriptional relationship (TaqMan mRNA expression assays (**d**); luciferase reporter assays (**e**)). **f**, Additional chromatin immunoprecipitation studies confirm that NF- κ B1 protein binds the CD47 promoter within 90 min of TNF- α -treatment in human coronary artery SMCs. ***P* < 0.01, **P* < 0.05. Error bars represent the s.e.m.



Extended Data Figure 9 | Dual inhibition of CD47 and TNF-α provides a combinatorial effect.

a, Pretreatment of mouse vascular SMCs with a chemical inhibitor (SPD 304) or a monoclonal antibody (infliximab) directed against TNF-α prevents the increase in CD47 expression normally seen after TNF-α exposure. **b**, Similar effects were observed with the NF-κB inhibitor, BAY 11-7085, confirming the molecular mechanism outlined in Fig. 4. **c**, Mice injected for four weeks with the decoy TNF-α receptor, etanercept, display a significant reduction in their *in vivo* expression of CD47 in splenic (left) and renal (right) tissue. **d**, Publicly available microarray data from human clinical trials of commercially available TNF-α inhibitors reveal that subjects treated with these agents also express lower levels of CD47 *in vivo* (as assessed by two-tailed *t*-tests), confirming the mouse findings above (GSE accession numbers from left to right: 16,879 (*n* = 85), 12,251 (*n* = 22), 47,751 (*n* = 28) and 41,663 (*n* = 66)). **e**, **f**, Additional *in vitro* efferocytosis assays confirm a synergistic effect of anti-CD47 antibodies with a variety of TNF-α inhibitors in both the absence (**e**) and presence (**f**) of exogenous TNF-α. **g**, Mice with established plaques of identical size and with equivalent rates of apoptosis were treated with a short course (5 days) of IgG, anti-CD47

antibodies, etanercept, or combination therapy before collection. As shown the phagocytic index (indicated by the ratio of 'free' (white stars) to 'associated' (white arrows) apoptotic bodies) displayed a non-significant trend towards improvement for combination therapy (*P* > 0.05).

h, When treated for a full 28 days in the angiotensin-infusion model, individual comparisons showed that etanercept alone had no effect on atherosclerosis, and combination therapy was not significantly different from anti-CD47 alone, probably due to the potent effect of anti-CD47 monotherapy. ANOVA post-hoc test analysis did identify a significant linear trend across all four groups (*P* for trend < 0.01).

i, Electron microscopy provides additional qualitative evidence that combination therapy may provide an incremental effect on efferocytosis, as suggested by an increased prevalence of macrophages within the plaque which had ingested a large number of apoptotic bodies (white arrows), a reduced prevalence of free apoptotic bodies (yellow arrows), and a reduced prevalence of uncleared cells undergoing secondary necrosis (red arrows). ****P* < 0.001, ***P* < 0.01, **P* < 0.05. Error bars represent the s.e.m.

Extended Data Table 1 | *In vivo* serological data and additional *in silico* and bioinformatic data

Characteristic	Reference Range	anti-CD47 Ab	IgG	P value	Bioinformatic algorithm	Molecule Type	P value of overlap	Bioinformatic algorithm	FDR	Fold Enrichment	Benfordson	P value
Complete Blood Count					Upstream Regulator INGENUITY Pathway Analysis			KEGG Pathway Analysis				
WBC	5.5–9.3 K/ μ L	1.55	1.58	0.50	TNF	cytokine	6.50E-42	CD47 pathway analysis	2.47E-11	8.49	1.08E-12	2.38E-01
RBC	7.0–8.9 M/ μ L	3.64	11.27	0.01	IL4	cytokine	7.31E-41	Cytokine-cytokine receptor interaction	3.00E-09	18.11	8.76E-11	1.27E-12
Hb	18.7–16.4 g/dL	13.91	16.15	0.01	TGF β 1	growth factor	1.97E-39	Interleukin network for lipid production	4.68E-05	17.34	6.06E-07	9.71E-01
HCT	39.0–47.0 %	46.02	50.97	0.08	IL6	cytokine	1.27E-38	Adiponectin metabolism (AAS)	2.20E-08	8.81	1.08E-05	2.28E-01
MCV	57.0–68.7 fL	88.02	45.26	0.00	IL2	cytokine	2.00E-38	Individual hormone network for lipid production	2.58E-06	8.81	1.08E-05	2.28E-01
MCN	18.4–19.6 g	14.31	14.37	0.77	IL10	cytokine	4.21E-37	Cytokine-cytokine receptor interaction	3.71E-08	16.62	2.78E-05	3.00E-01
MCPC	89.0–96.0 g/dL	79.98	11.72	0.00	IL12	cytokine	2.00E-38	T cell receptor signaling pathway	5.18E-01	8.81	3.76E-04	5.65E-01
Platelet count	675–1338 K/ μ L	793.60	827.35	0.63	IL13	cytokine	1.29E-33	Interleukin-10 signaling pathway	1.18E-02	9.85	7.98E-04	1.13E-01
Reticulocyte Count	1.0–2.8 %	18.76	4.97	0.00	IL15	cytokine	9.08E-23	Apoptosis	1.84E-06	8.52	9.18E-01	1.27E-01
Reti Absolute	K/ μ L	1764586.40	553836.36	0.00	IL17	cytokine	3.08E-33	Prostate cancer	1.52E-05	8.33	1.05E-02	1.46E-01
Liver Function Studies					IL18	cytokine	3.33E-32	Primary immunodeficiency	2.88E-05	15.13	1.91E-02	2.72E-01
AST	352–388 U/L	120.00	290.56	0.24	IL20	cytokine	3.33E-32	Tuberculosis	3.08E-05	7.34	2.99E-02	2.94E-01
ALT	76–180 U/L	31.57	39.11	0.57	IL21	cytokine	3.09E-31	Cell-cell adhesion	4.43E-05	13.58	2.91E-02	6.16E-01
Alkaline Phosphatase	171–183 K/L	121.43	135.44	0.54	IL22	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Total Protein	5.0–6.2 g/dL	5.90	5.30	0.29	IL23	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Albumin	3.2–3.6 g/dL	2.93	2.88	0.27	IL24	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Globulin	N/A	2.97	2.39	0.06	IL25	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
A/G Ratio	N/A	1.11	1.54	0.38	IL26	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Total Bilirubin	N/A mg/dL	0.29	0.19	0.27	IL27	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Liver Chemistries					IL28	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
BUN	20.3–24.7 mg/dL	36.86	35.89	0.83	IL29	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Creatinine	0.1–1.1 mg/dL	0.30	0.17	0.04	IL30	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
BUN/Cr	N/A	132.93	205.71	0.22	IL31	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Ca	8.8–9.7 mg/dL	7.81	8.53	0.46	IL32	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Na	135–145 mmol/L	146.67	145.25	0.60	IL33	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
K	3.5–5.0 mmol/L	3.47	3.20	0.80	IL34	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Cl	N/A mmol/L	112.67	107.00	0.08	IL35	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Carbon Dioxide	N/A mmol/L	17.60	18.54	0.38	IL36	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Ca	N/A	14.77	16.08	0.60	IL37	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Anion gap	N/A mEq/L	25.87	28.93	0.20	IL38	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Uric acid	184–220 mg/dL	116.71	109.44	0.86	IL39	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Metabolic Measurements					IL40	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Cholesterol	N/A mg/dL	508.50	554.20	0.31	IL41	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Triglycerides	N/A mg/dL	105.80	101.70	0.76	IL42	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
HDL	N/A mg/dL	391.40	508.80	0.22	IL43	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
LDL	N/A mg/dL	480.10	476.30	0.36	IL44	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Insulin	N/A μ g/L	0.93	0.81	0.81	IL45	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Lipid Panel					IL46	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
HDL	25–32%	27.83	32.25	0.64	IL47	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
LDL	65–80%	62.00	62.00	0.50	IL48	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Triglycerides	0–3%	4.17	4.00	0.88	IL49	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01
Insulin	0–3%	3.00	3.50	0.79	IL50	cytokine	3.33E-32	Immune system	4.43E-05	13.58	2.91E-02	6.16E-01

a, Complete serological studies (including blood count, liver function studies, basic metabolic panel and fasting glucose) from the 4-week *apoE^{-/-}*-AngII atherosclerosis model indicate that anti-CD47 antibody induces a significant reduction in haemoglobin and compensatory reticulocytosis, consistent with previous reports^{4,7}. The erythrophagocytosis of senescent red blood cells appears to be self-limited, and no anaemia was observed in the chronic atherosclerosis model or the reduced dose model ($P=0.54$ and 0.57 , respectively). No significant difference in any other serum marker is observed except for an increase in serum creatinine, which does not deviate outside of the reference range. Metabolic parameters and leukocyte differential data from the 12-week chronic atherosclerosis model are displayed at the bottom of the table. b, Additional Upstream Regulator Analysis (URA) bioinformatic analyses of the cytoscape data displayed in Extended Data Fig. 7a performed within the Ingenuity Pathway Analysis (IPA) software identifies a number of TNF- α -related factors (indicated in red) which are predicted to mediate transcriptional regulatory roles in the gene network shown in that panel. P values were determined by Fisher's exact test by comparing overlap of co-expressed genes with known upstream regulators from the Ingenuity Knowledge Base. c, Several additional DAVID-based bioinformatics analyses including (KEGG, SMART, PANTHER and GO analyses) confirm the association between CD47 and inflammatory signalling related to the TNF- α pathway (indicated in red). Blue panels indicate the $-\log_{10}(P)$ value for each identified factor. d, Transcription factor binding site prediction algorithms identify several putative NF- κ B family binding sites within the CD47 promoter, as displayed in Extended Data Fig. 8a. e, List of primers used in this study.

Neoantigen landscape dynamics during human melanoma–T cell interactions

Els M. E. Verdegaal¹, Noel F. C. C. de Miranda², Marten Visser¹, Tom Harryvan¹, Marit M. van Buuren^{3†}, Rikke S. Andersen^{4†}, Sine R. Hadrup^{4†}, Caroline E. van der Minne¹, Remko Schotte⁵, Hergen Spits^{5,6}, John B. A. G. Haanen^{1,3}, Ellen H. W. Kapiteijn¹, Ton N. Schumacher³ & Sjoerd H. van der Burg¹

Recognition of neoantigens that are formed as a consequence of DNA damage is likely to form a major driving force behind the clinical activity of cancer immunotherapies such as T-cell checkpoint blockade and adoptive T-cell therapy^{1–7}. Therefore, strategies to selectively enhance T-cell reactivity against genetically defined neoantigens^{1,8–11} are currently under development. In mouse models, T-cell pressure can sculpt the antigenicity of tumours, resulting in the emergence of tumours that lack defined mutant antigens^{12,13}. However, whether the T-cell-recognized neoantigen repertoire in human cancers is constant over time is unclear. Here we analyse the stability of neoantigen-specific T-cell responses and the antigens they recognize in two patients with stage IV melanoma treated by adoptive T-cell transfer. The T-cell-recognized neoantigens can be selectively lost from the tumour cell population, either by overall reduced expression of the genes or loss of the mutant alleles. Notably, loss of expression of T-cell-recognized neoantigens was accompanied by development of neoantigen-specific T-cell reactivity in tumour-infiltrating lymphocytes. These data demonstrate the dynamic interactions between cancer cells and T cells, which suggest that T cells mediate neoantigen immunoediting, and indicate that the therapeutic induction of broad neoantigen-specific T-cell responses should be used to avoid tumour resistance.

Cytotoxic CD8⁺ T cells can specifically recognize and eliminate tumours^{14–16}. The clinical activity of antibodies against the T-cell checkpoints CTLA-4 and PD-1 underscores the notion that in many patients a tumour-specific T-cell response is present¹⁷. Similarly, the clinical effects of autologous adoptive cell transfer (ACT) rely on the presence of tumour-reactive T-cell populations^{18–20}. Cancer genome-screening strategies to dissect T-cell responses in these treated patients provide strong evidence for a role of neoantigen recognition. Neoantigen-specific CD4⁺ and CD8⁺ T-cell responses are frequently observed in melanoma and gastrointestinal tract tumours^{3,5,11,21}. T-cell responses against neoantigens can be enhanced by CTLA-4 and PD-1 blockade^{4,7}, and mutational load correlates with response to checkpoint blockade in non-small cell lung cancer, melanoma, and tumours with mismatch repair deficiency^{2,4,6}. Finally, a recent case report provides direct evidence for the clinical potential of neoantigen-specific T-cell reactivity¹¹.

These data, and the fully tumour-restricted expression of neoantigens, makes them an attractive target for novel cancer immunotherapies. In support of this concept, neoantigen vaccines were effective in mouse tumour models and immunogenic when injected in humans^{1,8,9}. Neoantigen-specific T-cell transfer strategies are also being developed²².

However, loss of a defined neoantigen resulted in tumour cell resistance in a transplantable tumour model¹². Whether loss of

neoantigens also occurs in human disease is unclear. Therefore, we analysed the changes in expression and recognition of neoantigens targeted by tumour-specific T cells during the follow-up of two patients with stage IV melanoma treated with ACT. The first patient (patient BO) presented with stage IV melanoma with multiple (>9) subcutaneous, lymph node and lung lesions. One subcutaneous lesion was resected to establish melanoma cell line MEL05.18. One year later, the disease progressed and the patient developed a brain metastasis that was partially resected and used to establish both a melanoma line (MEL06.07) and tumour-infiltrating lymphocytes (TIL06.07). At this point, ACT treatment was initiated, using a tumour-specific T-cell product expanded by repeated stimulation of peripheral blood mononuclear cells (PBMC) with MEL05.18. This treatment led to a complete response that has been ongoing for more than 9 years (Fig. 1a).

The infusion product comprised 70% of CD8⁺ T cells. Stimulation with MEL05.18 showed that >80% of the CD8⁺ and >15% of the CD4⁺ T cells were tumour reactive²³. Analysis of CD8⁺ T-cell reactivity using a panel of >200 major histocompatibility complex (MHC) multimers containing known shared tumour-associated epitopes^{14,24} demonstrated that only a minor fraction of tumour-reactivity could be explained by recognition of these shared antigens (1.24% of the CD8⁺ T cells responded to three different gp100 epitopes; data not shown). The dominant recognition of unique patient-specific (private) antigens was also suggested by stimulation of the T cells used for infusion with a panel of (partially) human leukocyte antigen (HLA)-matched melanoma lines, showing exclusive recognition of the autologous tumour (Fig. 1b).

To identify the antigens recognized, we analysed MEL05.18 for non-synonymous somatic mutations within expressed genes and found that MEL05.18 expressed 501 non-synonymous mutated genes. Autologous antigen-presenting cells²⁵ loaded with synthetic 31-mer peptides covering these mutations were used to assess T-cell recognition.

In the infusion product used to treat patient BO, we observed CD4⁺ T-cell responses against RPS12(V104I), ZC3H18(G269R) (Fig. 1c and ref. 3) and TNK(S502F)³, and CD8⁺ T-cell responses against KIAA0020(P451L) (KIA(P451L)) and ribosomal protein RPL28(S76F) (Fig. 1d). The combined frequency of CD8⁺ T cells that recognize these two mutant epitopes equalled the total tumour-reactive CD8⁺ T cell pool in the infusion product (Fig. 1d and 1e).

To address the stability of the neoantigen repertoire recognized by CD8⁺ T cells, we compared neoantigen reactivity in peripheral blood and resected tumour tissue. CD8⁺ T cell responses against both of the identified neoepitopes were detectable *ex vivo* in the blood of patient BO that was sampled before ACT. Their frequency increased after ACT (Fig. 1f) and eventually contracted following complete tumour

¹Department of Medical Oncology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ²Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ³Department of Immunology, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. ⁴Department of Hematology, Center for Cancer Immune Therapy, University Hospital Herlev, 2730 Herlev, Denmark. ⁵AIMM Therapeutics, 1105 BA Amsterdam, The Netherlands. ⁶Department of Cell Biology and Histology, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. †Present addresses: Neon Therapeutics, Cambridge, Massachusetts 02142, USA (M.M.v.B.); Department of Cancer and Inflammation Research, Institute for Molecular Medicine, University of Southern Denmark, 5000 Odense, Denmark (R.S.A.); Section for Immunology and Vaccinology, National Veterinary Institute, Technical University of Denmark, 1870 Frederiksberg, Copenhagen, Denmark (S.R.H.).

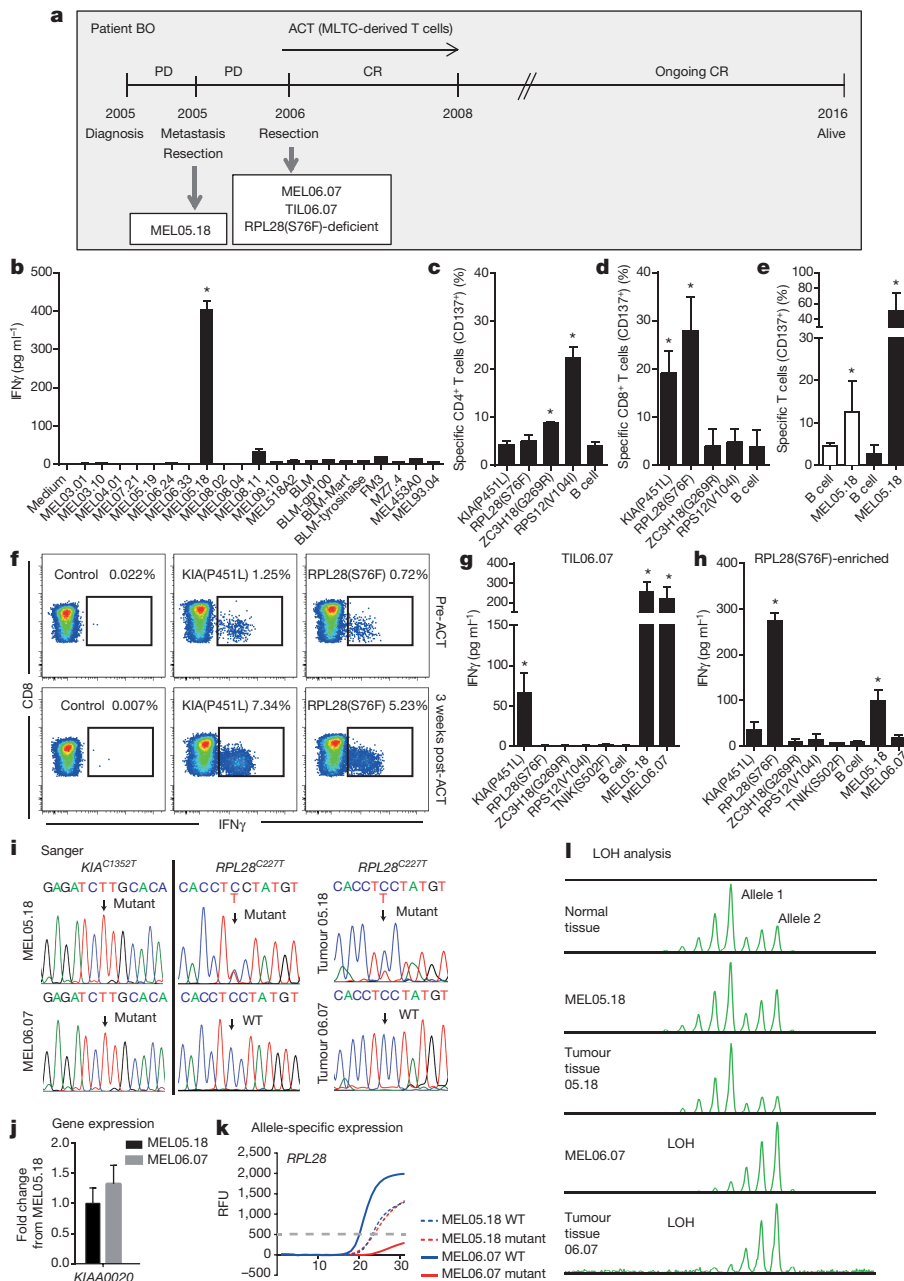


Figure 1 | Identification and dynamics of neoepitope-specific T-cell reactivity and neoepitope expression in patient BO. **a**, Clinical course, tissue and cell line collection. CR, complete response; PD, progressive disease. **b**, T-cell reactivity against autologous (asterisk) and a panel of (partially) HLA class I and II matched melanoma cell lines by IFN γ release. The mean \pm s.d. of a representative experiment performed in duplicate is depicted. **c–e**, Neoantigen-specific CD137 $^{+}$ cell-frequency among CD4 $^{+}$ (**c**) or CD8 $^{+}$ T cells (**d**). Autologous tumour cells and B cells served as positive and negative controls for CD4 $^{+}$ (white bars) and CD8 $^{+}$ (black bars) T cells (**e**). The mean \pm s.d. of at least two independent experiments is shown. Asterisk indicates significant increase above negative control ($P < 0.05$; Student's t -test). **f**, Percentage of IFN γ -producing KIA(P451L)-specific (middle) and RPL28(S76F)-specific (right) CD8 $^{+}$ cells in peripheral blood before and 3 weeks after ACT.

regression (not shown). Surprisingly, although the TIL06.07 culture isolated from the brain tumour resected just before ACT comprised almost exclusively CD8 $^{+}$ T cells, T-cell reactivity was only observed against the KIA(P451L) neoantigen, and not against RPL28(S76F) (Fig. 1g). Whereas the mutation and expression of the *KIA0020* gene (also known as *PUM3*) was present in both cell lines (Fig. 1i and j),

g, TIL06.07 reactivity against the identified neoantigens, MEL05.18 and MEL06.07. **h**, Reactivity of enriched RPL28(S76F)-specific T cells against neoantigens, MEL05.18 and MEL06.07. Autologous B cells served as negative controls (**f–h**). Mean \pm s.d. of a representative experiment performed in triplicate is shown (**g–h**). Asterisks indicate significant increase above negative control ($P < 0.01$; Student's t -test). **i**, Sanger sequencing of the indicated genes in cell lines and tissue. The arrows indicate the non-synonymous mutations resulting in the identified neoantigens. WT, wild type. **j**, Fold change in RNA expression of *KIA0020* in MEL06.07 relative to MEL05.18, mean \pm s.d. of two independent experiments is shown. **k**, Allele-specific expression of the wild-type (blue) and mutant (red) allele of the *RPL28* gene. Grey dashed line depicts cut-off value for specific expression. **l**, Loss of heterozygosity (LOH) near the *RPL28* gene in both MEL06.07 and tumour tissue 06.07.

absence of RPL28(S76F)-specific T cells coincided with absence of the mutant allele encoding the RPL28(S76F) neoantigen within the MEL06.07 cell line (Fig. 1i and k). To assess if the cell lines were representative of the corresponding tumour samples, Sanger sequencing of the *RPL28* mutation was performed in archival formalin-fixed tissues and was consistent throughout (Fig. 1i). Loss of heterozygosity (LOH)

analysis by amplification of polymorphic markers around the *RPL28* gene confirmed loss of the mutant *RPL28* allele in tumour tissue 06.07, whereas both alleles were present in early tumour tissue 05.18 and normal tissue (Fig. 1l). The presence of a single detectable *RPL28* allele in 06.07 tumour tissue indicates the clonal outgrowth of a tumour subclone with LOH at this locus, and is consistent with absence of the *RPL28*(S76F) mutation within the MEL06.07 cell line and 06.07 tissue (Fig. 1i). In line with these findings, T cells enriched for *RPL28*(S76F) reactivity responded to MEL05.18 but not to the *RPL28*(S76F) loss variant MEL06.07 (Fig. 1h).

The loss of a neoantigen recognized by T cells in a sequential tumour lesion is consistent with the concept of immunoediting. Furthermore, the data suggest that accumulation of neoantigen-specific T cells at the tumour site reflects local antigen presence. To address this, we investigated neoantigen expression and reactivity in sequential lesions from a second patient. Patient AB was surgically treated for metastatic melanoma that spread to the axillary lymph nodes. From the resected tumour tissue, melanoma line MEL04.01 was established and used to expand tumour-specific T cells from the peripheral blood. Two years later, the patient developed a progressive liver lesion that was treated by ACT, resulting in disease stabilization for 3 months according to response evaluation criteria in solid tumours (RECIST1.1) and 6 months according to immune-related response criteria (iRC)²⁶. Subsequently, the liver lesion was resected and used to generate TIL08.10. The patient remained disease-free for a period of 4 years and then developed multiple subcutaneous, lung and bone lesions. Two subcutaneous lesions, 12.07 and 12.09, were resected for the isolation of TIL12.07 and TIL12.09, respectively, and to establish melanoma line MEL12.07 (Fig. 2a).

The infusion product of patient AB comprised 70% of CD8⁺ T cells, and 54% of these CD8⁺ T cells were reactive against MEL04.01 used to generate the infusion product, whereas no reactivity was observed within the CD4⁺ T-cell fraction²³. Again, dominant recognition of private antigens by infused T cells was suggested by the limited CD8⁺ T-cell reactivity against shared antigens^{14,24} (<0.04% of CD8⁺ T cells infused into patient AB recognized a shared PRDX5 or EPHA2 epitope; data not shown) and exclusive recognition of the autologous tumour within a panel of (partially) HLA-matched melanoma lines (Fig. 2b).

MEL04.01 of patient AB displayed 226 expressed non-synonymous mutations and the infusion product used for ACT comprised CD8⁺ T cells specific for the neoantigens EML1(R64W), SEPT2(R300C), and CAD(R1854Q) (Fig. 2c). In all cases, no or only weak recognition of the corresponding wild-type peptides was observed (data not shown).

Independent TIL cultures from tumour lesions of patient AB, resected before and after ACT, were screened for recognition of all putative neoepitopes present within the original MEL04.01 tumour (Fig. 2d–g). Interestingly, TILs obtained before ACT (TIL04.01) did not react against any of the three identified neoantigens and showed only weak reactivity against the MEL04.01 tumour from which they were derived. To assess the fate of the infused neoantigen-specific T-cell populations and the antigens they recognize, TIL08.10 obtained two years after therapy, and TIL12.07 and TIL12.09 obtained six years after therapy were analysed. A low-level T-cell response against the EML(R64W) neoantigen was detectable in all three TIL preparations, although not significantly above background in TIL12.07. Notably, within the 12.07 and 12.09 TIL preparations, no significant T-cell reactivity was detectable against the CAD(R1854Q) neoantigen, and reactivity against the SEPT2(R300C) neoantigen was absent from all post-ACT TIL preparations (Fig. 2e–g). Thus, in this patient immunity towards defined CD8⁺ T-cell neoantigens was observed after ACT, which was not detected in TIL before therapy, and which was again largely absent from subsequent lesions.

To address whether the observed loss of defined neoantigen-specific T cells over time could be compensated by acquisition of novel T-cell responses, TIL08.10, TIL12.07 and TIL12.09 were analysed for recognition of additional neoantigens that were encoded by mutations

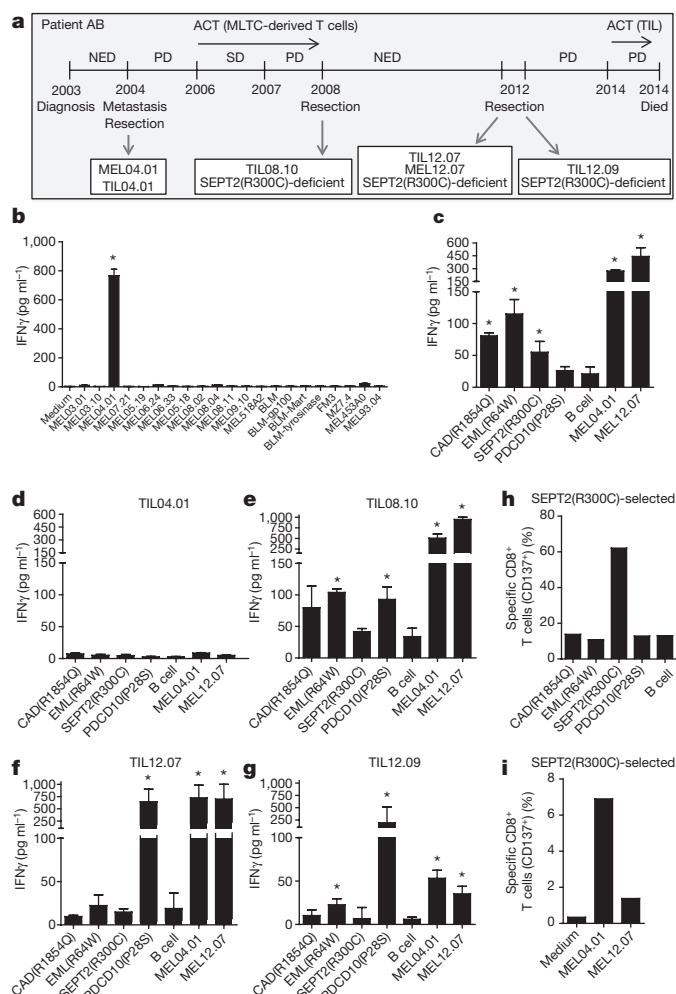


Figure 2 | Identification and dynamics of the dominating intratumoral neoantigen-specific T-cell repertoire in patient AB. **a**, Clinical course, tissue and cell line collection. **b–g**, T-cell reactivity measured by IFN γ release. **b**, Reactivity of ACT product against autologous (asterisk) and a panel of (partially) HLA class I and II matched melanoma cell lines. Mean \pm s.d. of a representative experiment performed in duplicate is depicted. **c**, Reactivity of ACT product against indicated neoepitope-loaded autologous B cells, autologous tumour cells, and unloaded autologous B cells (negative control). **d–g**, IFN γ release by TIL04.01 obtained before ACT (**d**), TIL08.10 obtained 2 years after ACT (**e**) and TIL12.07 (**f**) and TIL12.09 (**g**) at disease progression, stimulated with the indicated neoantigen-loaded B cells, MEL04.01 and MEL12.07. The mean \pm s.d. of at least two independent experiments is shown and significant increase above unloaded B cells as negative control ($P < 0.05$; Student's *t*-test) is depicted by an asterisk (**c–g**). **h**, **i**, T cells selected for SEPT2(R300C) reactivity specifically recognize SEPT2(R300C) and MEL04.01 but not MEL12.07, as indicated by the percentage of CD137⁺CD8⁺ cells. Unloaded B cells (**h**) or medium (**i**) served as background controls. A representative of two independent experiments is shown.

present in the original tumour MEL04.01. This revealed a low-level T-cell response towards the programmed cell death protein 10 (PDCD10(P28S)) neoantigen in TIL08.10 (Fig. 2e) and a pronounced T-cell response towards this neoantigen in both TIL12.07 and TIL12.09 (Fig. 2f and g, respectively).

To assess whether absence of detectable SEPT2(R300C), and CAD(R1854Q) neoantigen-specific T-cell reactivity in TIL derived from subsequently resected post-ACT lesions coincided with altered neoantigen expression, we compared the cell lines derived from the original MEL04.01 tumour and the recurrent lesion MEL12.07 isolated eight years later. HLA genotyping and flow cytometry revealed that

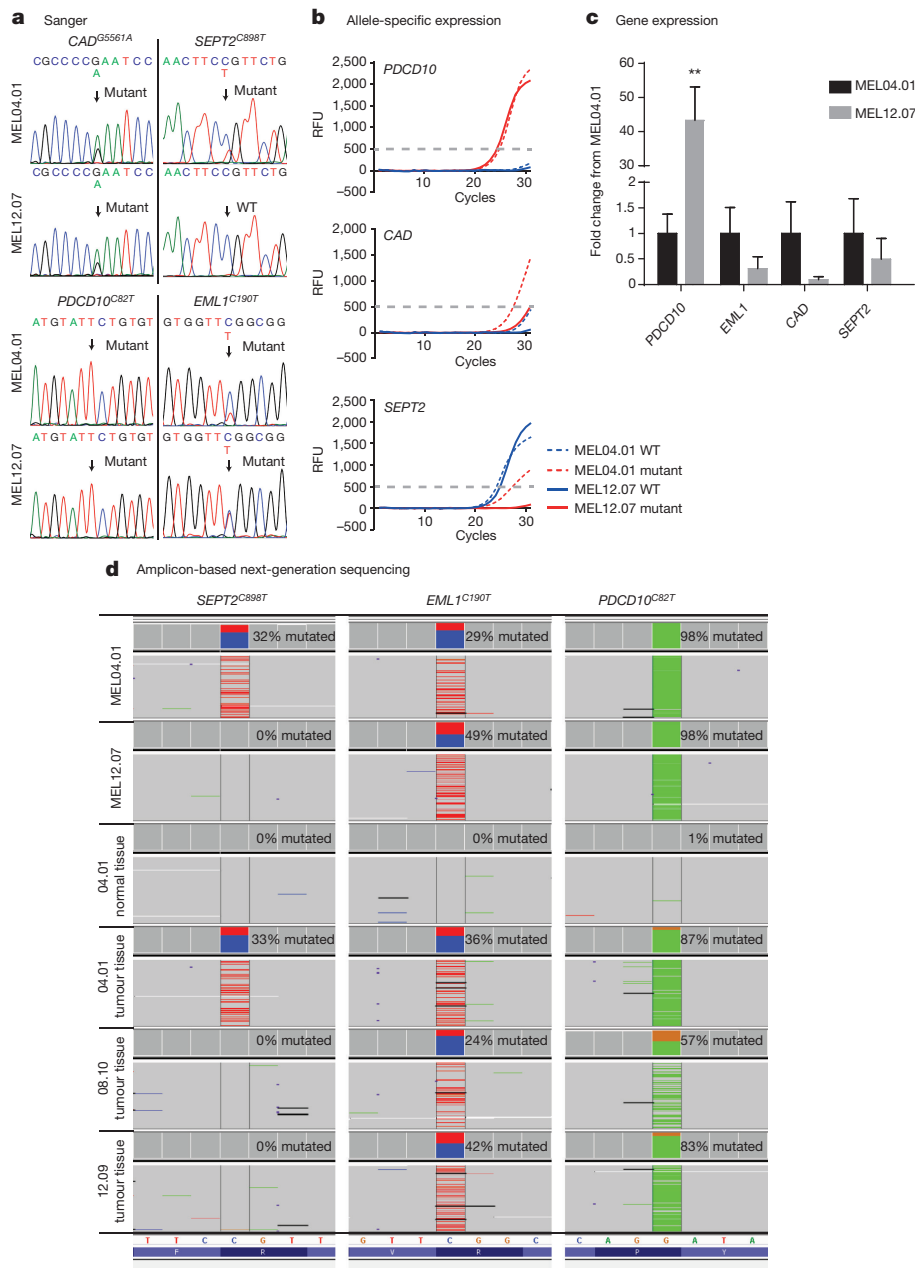


Figure 3 | Mutation analysis and neoantigen expression in sequential tumour lines and tumour tissue of patient AB. DNA sequencing (Sanger), allele-specific expression (KASP analysis) and RNA expression (qPCR) of cell lines derived from patient AB (MEL04.01 and MEL12.07). **a**, The arrows indicate the non-synonymous mutations resulting in the identified neoantigens. **b**, Allele-specific expression of the wild-type (blue) or mutant (red) alleles. Grey dashed line depicts cut-off value for specific

expression. **c**, The fold change in RNA expression of the indicated genes in MEL12.07 versus MEL04.01. The mean \pm s.d. of three independent experiments is shown. **d**, Detection of *SEPT2*^{C898T} mutation in MEL04.01 and respective tumour sample but not in normal tissue or tumour samples 08.10 or 12.09. Collapsed sequencing reads are displayed. In grey, nucleotide positions correspond to the reference sequence. Mutations are displayed in red (>T) and green (>A).

both cell lines showed comparable HLA cell-surface expression and no loss of HLA alleles (data not shown). Notably, whereas DNA sequencing revealed that the mutation in the *CAD* gene was still present in cells derived from the recurring lesion (Fig. 3a), RNA expression analysis demonstrated an approximately tenfold reduction in expression of this gene (Fig. 3c). Furthermore, whereas RNA expression of the *SEPT2* gene was not significantly altered in the cell line derived from the recurring lesion (not shown), the mutant allele encoding the SEPT2(R300C) neoantigen that was present in the original tumour, was selectively lost in the tumour cell line derived from tumour tissue isolated eight years later (Fig. 3a, b). In agreement with these findings, selected SEPT2(R300C)-specific T cells (Fig. 2h) responded to MEL04.01 but not to SEPT2(R300C)-negative MEL12.07 tumour

cells (Fig. 2i). Next-generation amplicon sequencing revealed that the *SEPT2*^{C898T} mutation was present in approximately 30% of the sequencing reads spanning the mutation in both the 04.01 tumour tissue and its corresponding cell line, whereas it was already lost in the liver lesion 08.10 resected after ACT, as well as in lesion 12.09 (Fig. 3d). Of note, the *PDCD10*^{C82T} and the *EML1*^{C190T} mutations were also detected in the 04.01 tumour tissue at frequencies (87% and 36% of sequencing reads, respectively) similar to those observed within the 04.01 cell line (Fig. 3d). Furthermore, in contrast to the *SEPT2*^{C898T} mutation, these mutations were detected in a similar proportion of sequencing reads in the subsequent clinical samples (Fig. 3d), confirming that the different tumour lines were representative for the mutation status in the corresponding clinical samples. Notably, whereas the mutant

PDCD10 allele was already present in the first tumour line (MEL04.01), RNA expression of *PDCD10* had increased >40-fold in the recurring MEL12.07 lesion (Fig. 3c). Competitive allele-specific PCR (KASP) analysis confirmed that all *PDCD10* expression was derived from the mutant allele (Fig. 3b). Thus, the observed loss of a mutant allele or changes of RNA expression in the tumours of patient AB corresponded exactly with the observed changes in neoantigen-specific T-cell reactivity.

In depth analysis of the tumour-specific reactivity of PBMC, infusion products and TILs revealed neoantigen-specific T-cell reactivity in both long-term survivors after ACT. All mutations yielding these T-cell-recognized neoantigens were absent in the COSMIC cancer gene census except for SEPT2(R300C) that was described once²⁷, suggesting that these are passenger mutations that do not contribute to cellular fitness. Over time, we observed changes in expression of four out of six detected neoantigens. In two cases, the mutant allele was lost from a subsequent tumour, in one case expression of the mutant gene was substantially reduced, and in one case expression of the mutant gene was substantially increased. In four out of four cases, these changes in neoantigen expression were paralleled by loss or gain of neoantigen-specific T-cell reactivity. The current data may be explained both by random clonal variation between tumour metastases and by true T-cell-mediated selection of antigen-negative variants. Although we cannot distinguish between these two hypotheses, support for the latter interpretation is provided by the selective recognition of the early but not recurring lesions by both RPL28(S76F)-specific and SEPT2(R300C)-specific T cells of the two patients. Regardless, the current data demonstrate that the interaction between human melanoma and T cells varies over time, in a way that is similar to the interaction between the T-cell compartment and viral quasispecies²⁸. We also note that the genetic differences and the differences in RNA expression between the sequentially obtained tumours characterized here reflect heterogeneity that has been observed within many human tumours²⁹. As such, it is plausible that a similar drift of the neoantigen repertoire will also be a factor in other human malignancies. Collectively, the data presented here demonstrate that under conditions in which a high frequency of tumour-specific T cells is present, tumour-cell variants with reduced or lost neoantigen expression can emerge, similar to what has been observed in mice¹. The observation of concurrent acquisition of novel T-cell reactivity implies that cancer immunotherapies should aim to exploit the adaptive capacity of the immune system in order to maintain strong immune surveillance over time.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February; accepted 21 June 2016.

Published online 27 June 2016.

- Gubin, M. M. *et al.* Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
- Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
- Linnemann, C. *et al.* High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4⁺ T cells in human melanoma. *Nat. Med.* **21**, 81–85 (2015).
- Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
- Robbins, P. F. *et al.* Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* **19**, 747–752 (2013).
- Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
- van Rooij, N. *et al.* Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J. Clin. Oncol.* **31**, e439–e442 (2013).
- Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).

- Kreiter, S. *et al.* Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **520**, 692–696 (2015).
- Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
- Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4⁺ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
- Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoeediting. *Nature* **482**, 400–404 (2012).
- Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoeediting and its three component phases—elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27**, 16–25 (2014).
- Andersen, R. S. *et al.* Dissection of T-cell antigen specificity in human melanoma. *Cancer Res.* **72**, 1642–1650 (2012).
- Kawakami, Y. *et al.* Recognition of shared melanoma antigens in association with major HLA-A alleles by tumor-infiltrating T lymphocytes from 123 patients with melanoma. *J. Immunother.* **23**, 17–27 (2000).
- Lu, Y. C. *et al.* Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clin. Cancer Res.* **20**, 3401–3410 (2014).
- Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science* **348**, 56–61 (2015).
- Besser, M. J. *et al.* Clinical responses in a phase II study using adoptive transfer of short-term cultured tumor-infiltrating lymphocytes in metastatic melanoma patients. *Clin. Cancer Res.* **16**, 2646–2655 (2010).
- Dudley, M. E. *et al.* Adoptive cell therapy for patients with metastatic melanoma: evaluation of intensive myeloablative chemoradiation preparative regimens. *J. Clin. Oncol.* **26**, 5233–5239 (2008).
- Ellebaek, E. *et al.* Adoptive cell therapy with autologous tumor-infiltrating lymphocytes and low-dose interleukin-2 in metastatic melanoma patients. *J. Transl. Med.* **10**, 169–180 (2012).
- Tran, E. *et al.* Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science* **350**, 1387–1390 (2015).
- Gros, A. *et al.* PD-1 identifies the patient-specific CD8⁺ tumor-reactive repertoire infiltrating human tumors. *J. Clin. Invest.* **124**, 2246–2259 (2014).
- Verdegaal, E. M. *et al.* Successful treatment of metastatic melanoma by adoptive transfer of blood-derived polyclonal tumor-specific CD4⁺ and CD8⁺ T cells in combination with low-dose interferon- α . *Cancer Immunol. Immunother.* **60**, 953–963 (2011).
- Hadrup, S. R. *et al.* Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat. Methods* **6**, 520–526 (2009).
- Kwakkenbos, M. J. *et al.* Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. *Nat. Med.* **16**, 123–128 (2010).
- Wolchok, J. D. *et al.* Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin. Cancer Res.* **15**, 7412–7420 (2009).
- Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic *RAC1* mutations in melanoma. *Nat. Genet.* **44**, 1006–1014 (2012).
- Timm, J. & Walker, C. M. Mutational escape of CD8⁺ T cell epitopes: implications for prevention and therapy of persistent hepatitis virus infections. *Med. Microbiol. Immunol.* **204**, 29–38 (2015).
- Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).

Acknowledgements We are grateful to G. J. Liefers for handling of patient material, and T. van Wezel and D. Ruano for the setup of the M13-amplicon sequencing technology. This work was supported by Dutch Cancer Society grant UL 2012-5544 (to E.M.E.V., S.H.v.d.B. and J.B.A.G.H.), the Anticancer Fund (to E.M.E.V. and S.H.v.d.B.), Dutch Cancer Society grant NKI 2012-5463 (to T.N.S., J.B.A.G.H. and S.H.v.d.B.), the Dutch Cancer Society Queen Wilhelmina Award NKI 2013-6122 (T.N.S.), Dutch Cancer Society grant UVA 2010-4822 (to H.S.), Fight Colorectal Cancer-Michael's Mission-AACR Fellowship (2015) and Alpe d'HuZes/KWF Bas Mulder Award (to N.F.C.C.d.M.).

Author Contributions E.M.E.V. designed, performed, analysed and interpreted experiments and wrote the paper, N.F.C.C.d.M., M.V., C.E.v.d.M. and T.H. designed, performed, analysed and interpreted the experiments, M.M.v.B. analysed and interpreted next-generation sequencing data and performed peptide selection, R.S. and S.R.H. designed, performed and interpreted the combinatorial coding experiments, R.S. generated BCL-6/BCL-XL immortalized B-cell lines, E.H.W.K. and J.B.A.G.H. supervised treatment of patients, supplied patient material and provided clinical interpretation of results, H.S. developed the BCL-6/BCL-XL immortalization technology, T.N.S. interpreted the data and wrote the paper, S.H.v.d.B. supervised the project, designed and interpreted the experiments, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Correspondence and requests for materials should be addressed to E.M.E.V. (E.Verdegaal@lumc.nl).

Reviewer Information Nature thanks M. Gubin, U. Sahin and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Data reporting. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size.

Subjects. The study was approved by the Medical Ethical Committee of the Leiden University Medical Center and written informed consent was obtained from all patients.

Cell lines and culture. Melanoma cell lines were previously generated from resected lesions of patients 05.18 and 04.01 (ref. 23). All other melanoma cell lines were also established in the laboratory of Medical Oncology (LUMC, Leiden, Netherlands) except for melanoma cell line BLM that was obtained from the Netherlands Cancer Institute (Amsterdam, Netherlands), FM3 and FM6 were provided by P. thor Straten and MZ7.4-mel was obtained from Johannes Gutenberg University. Cell lines were authenticated by HLA-genotyping, and were mycoplasma negative. All melanoma cell lines were cultured in tumour cell medium (that is, DMEM (Life Technologies, Breda, The Netherlands) with 8% heat-inactivated fetal calf serum (FCS), penicillin (100 IU ml⁻¹), streptomycin (100 µg ml⁻¹) and L-glutamine (4 mM) all from Lonza Biowhittaker (Breda, Netherlands)). HLA genotyping of the cell lines was performed by the Department of Immunohematology and Blood Bank of the LUMC. Autologous immortalized B cells, obtained by transduction with the anti-apoptotic genes Bcl-6 and Bcl-XL²⁵ were provided by AIMM therapeutics (Amsterdam, The Netherlands) and were cultured in IMDM (Lonza) with 8% heat-inactivated FCS, penicillin (100 IU ml⁻¹), streptomycin (100 µg ml⁻¹) and L-glutamine (4 mM). B cells were stimulated twice a week with irradiated (70 Gy) CD40L expressing mouse L-cell fibroblasts and murine IL-21 (50 ng ml⁻¹; AIMM therapeutics).

TIL were obtained by culturing small tumour fragments in T-cell medium (IMDM with 7.5% heat-inactivated pooled human serum (Sanquin Blood bank, Dordrecht, The Netherlands), penicillin (100 IU ml⁻¹), streptomycin (100 µg ml⁻¹) and L-glutamine (4 mM) supplemented with 1,000 IU ml⁻¹ recombinant human (rh) IL-2 (Aldesleukin, Novartis). TIL medium plus rhIL-2 was refreshed every two to three days. Tumour-specific T-cell batches were obtained by MLTC as previously described²³. Briefly, peripheral blood mononuclear cells were incubated with irradiated (100 Gy) autologous tumour cells at a 15:1 effector:target cell ratio in T-cell medium supplemented with rhIL-4 (10 U ml⁻¹; Cellgenix, Cellgro) to prevent expansion of natural killer (NK) cells. From day 2 onward, medium was refreshed every two to three days with T-cell medium plus rhIL-2 (150 IU ml⁻¹). T cells were re-stimulated weekly with irradiated autologous tumour cells and cultured for a total of 4 weeks before cryopreservation for further experiments and ACT treatment²³.

To test for reactivity of T cells used for ACT against shared antigens a panel of (partially) HLA-matched melanoma cell lines was used. At least one cell line with a matched allele for every HLA class II allele was tested. For HLA class I alleles, 3–8 matched cell lines are tested for each allele, except for the heterozygously expressed HLA-A*23.01 and HLA-B*49.01 in patient BO, for which no matched cell lines were available.

Neoantigen identification. Whole-exome and RNA sequencing was performed and 31-mer synthetic peptides covering the non-synonymous somatic mutations were manufactured as previously described³. Importantly, no selection based on *in silico* prediction of MHC binding affinity was used in order not to exclude any potential epitopes preemptively. Next, T cells were incubation with target cells that is, tumour cells or Bcl-6/Bcl-XL B cells either unloaded or preloaded overnight with peptide pools or single peptides (10 µg ml⁻¹ per peptide). Reactivity of T cells was measured after 24–48 h co-incubation with target cells by IFN γ secretion using ELISA (Sanquin) or intracellular cytokine staining for IFN γ or expression of the activation marker CD137 by FACS analysis, as previously described²³. Where indicated, neoepitope-specific T cells were selected after overnight stimulation of T cells with specific Bcl-6/Bcl-XL B cells loaded with synthetic 31-mer peptides using the CD137 MicroBead kit (Miltenyi Biotec) according to the manufacturer's instructions.

Sanger sequencing of DNA samples. Cell line DNA was extracted with the Genomic Wizard DNA purification kit (Promega). DNA from microdissected formalin-fixed archival material was extracted with the NucleoSpin FFPE DNA kit (Macherey-Nagel). Primers were designed with Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (see below). PCR reactions were carried out in 15 µl with 10 ng of DNA, 1 × iQ Supermix (Bio-rad), and 10 pmol of forward and reverse primer. PCR products were purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and sequenced in both forward and reverse directions at Macrogen Europe (Meibergdreef).

Real-time PCR gene expression analysis. Total RNA was isolated with the TRIzol (ThermoFisher) method including DNase treatment in suspension using rDNase (Macherey-Nagel). cDNA was synthesized using 2 µg of RNA, 50 ng of p(dT)15 (Roche), 0.3 µg of Random Primers (ThermoFisher), 1 mM dNTPs

(ThermoFisher), 4 µl 5 × AMV-RT buffer (Roche), 10 units of AMV-RT (Roche) and 20 units of RNasin Ribonuclease Inhibitor (Promega). cDNA synthesis was carried out at 42 °C for 1 h. qPCR primers were designed with Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (see below).

qPCR mixes consisted of: 2 µl of 1:25 cDNA solution, 3 pmol of primer (forward and reverse), iQ SYBR Green Supermix in 1 µl for *PDCD10* and *CPSF6* (housekeeping gene); 5 µl of 1:25 cDNA solution, 6 pmol of primer (forward and reverse), iQ SYBR Green Supermix (Bio-rad). Amplification cycles: 95 °C, 10 s; 60 °C, 30 s for *CPSF6* and *PDCD10*. 95 °C, 10 s; 62 °C, 30 s; 72 °C, 30 s for *EMLI1*, *CAD*, and *SEPT2*. Melting was performed to confirm the specificity of the assays. The ΔC_t method was applied to calculate the levels of gene expression, relative to the housekeeping gene. At least two independent measurements were performed to assess gene expression and the fold change in expression in the recurrent cell line relative to expression in the original cell line was calculated.

Amplicon next-generation sequencing. Amplicon next-generation sequencing was used to confirm that the cell lines were representative for the corresponding tumour tissue. Two independent primer sets per target were designed (see below) to amplify the *SEPT2*^{C898T}, *PDCD10*^{C82T}, and the *EMLI1*^{C190T} mutations in MEL04.01, MEL12.07, and their respective clinical samples. 18-mer, M13 tails were added to the 5'-end of the forward (5'-TGTAACACGACGCCAGT-3') and reverse (5'-CAGGAAACAGCTATGACC-3') primers. Two multiplex PCRs were performed for each sample; each PCR reaction containing one of the primer sets for each target. PCRs were performed with the FastStart High Fidelity PCR System (Roche). Following the pooling of both reactions and purification with AMPure XP beads (Beckman Coulter), a second PCR was performed in order to integrate A and P1 adaptor sequences for Ion Torrent and distinct barcodes for each sample. Sample libraries were quantified by qPCR, normalized and analysed by the Ion PGM system in an Ion 318 chip (ThermoFisher Scientific). More than 10,000 sequencing reads covering each mutation were generated.

Allele-specific expression. Single nucleotide polymorphism (SNP) genotyping was performed using the competitive allele-specific PCR (KASPar) assay, following the manufacturer's protocol (LGC Genomics). The oligonucleotides were designed using Primerpickr (KBioscience) (see below).

After initial denaturation at 94 °C for 15 min, 10 cycles consisting of 94 °C for 20 s and touchdown from 65 °C in 0.8 °C steps per cycle were run. Subsequently, 30 cycles of 94 °C for 20 s, 57 °C for 1 min, and 25 °C for 10 s followed by fluorescent detection were applied. Three independent measurements were performed to assess allele-specific expression. The cut-off value for specific expression is set at >500 relative fluorescence units (RFU) based on the corresponding positive control.

Loss of heterozygosity. The presence of polymorphic markers around the *RPL28* and *SEPT2* genes was investigated with the UCSC genome browser. *AFMA357YH1* and *DI9S1142*, and *AFM182YA5* and *RH56150*, were employed to determine the zygosity status around the *RPL28* and *SEPT2* genes, respectively. Primers were designed around these markers (see below) and hexachloro-fluorescein label was added to the 5'-end of the forward primer. PCR amplifications were performed with 10 ng of DNA in iQ Supermix (Bio-Rad). PCR products were mixed in a formamide solution containing GeneScan 500 TAMRA Size Standard (Applied Biosystems). Thereafter, samples were loaded in a 4-capillary 3130 DNA Analyzer and results were interpreted with the GeneMapper 4.1 software (Applied Biosystems). The only marker that was informative, that is, that presented a heterozygous status in the patient's germline DNA was *AFMA357YH1*.

Used primer sequences. For PCR and Sanger sequencing, the following primers were used: *KIAA0020* forward 5'-GGATGGGTAACTGGACTTCTGG-3', *KIAA0020* reverse 5'-GAAAGGTCCTGTGGTACTTTGTT-3'; *RPL28* forward 5'-GGGC TATGAGTGTGGCAGAAG-3', *RPL28* reverse 5'-AAAGCAAGAATC CATCCCTCTC-3'; *CAD* forward 5'-CCCTCCAGACACCTGAAAGAC-3', *CAD* reverse 5'-GGGCATGTGAGAGAGCTGTGA-3'; *SEPT2* forward 5'-AGA CATTTCCACGGCCATACT-3', *SEPT2* reverse 5'-GCTCAGGGTGGA CACAATACAGA-3'; *PDCD10* forward 5'-TGCCTAACGCACCGATAAGA-3', *PDCD10* reverse 5'-TGTTCTTTCTTCTCTTTCCACCA-3'; *EMLI1* forward 5'-TGCAAGTAGCATGGAGGTGA-3', *EMLI1* reverse 5'-CAAGATT CTGCCACGAGACCA-3'.

RT-PCR primers: *KIAA0020* forward 5'-CTGGAGACGTTACGCTACC-3', *KIAA0020* reverse 5'-TGTGAAGCTCTCCGTCCTTG-3'; *PDCD10* forward 5'-ACCGCAGGGCACTTGAAC-3', *PDCD10* reverse 5'-GGTTGGC ACTTACGAACACA-3'; *EMLI1* forward 5'-TGGGTTCCCTCTGCCTGTAA-3', *EMLI1* reverse 5'-CCTTCTGGCCACACTCCAAA-3'; *CAD* forward 5'-GAGATGACCACGACACCTGAA-3', *CAD* reverse 5'-GCTCCTCAG CTGGCAAA-3'.

KASPar primers. *RPL28* forward A1 5'-GAAGGTGACCAAGTTC ATGCTCAGCGGAAGCCTGCCACCTT-3', *RPL28* forward A2 5'-GAAGGTGCGAGTCAACGGATTAGCGGAAGCCTGCCACCTC-3', *RPL28*

reverse 5'-GCGTGGCGCGAGCATTCTTGT-3'; *PDCD10* forward A1 5'-GAAGGTGACCAAGTTCATGCTACTCGTTCTAGCTCATTAAACACAGA-3', *PDCD10* forward A2 5'-GAAGGTGCGAGTCAACGGATTCTCGTTCTAGCTCATTAAACACAGG-3', *PDCD10* reverse 5'-CTATGCCCCTCTATGCAGTCATGTA-3'; *CAD* forward A1 5'-GAAGGTGACCAAGTTCATGCTGGTCGGAGGCTCGATGGATT-3', *CAD* forward A2 5'-GAAGGTGCGAGTCAACGGATTGGTTCGGAGGCTCGATGGATT-3', *CAD* reverse 5'-GCTTCCTGATGGCCGCTTCCAT-3'; *SEPT2* forward A1 5'-GAAGGTGACCAAGTTCATGCTCCTCTCTTGAGTCTCTCAGAAC-3', *SEPT2* forward A2 5'-GAAGGTGCGAGTCAACGGATTCTCTCTTGAGTCTCTCAGAAC-3', *SEPT2* reverse 5'-GGTGACCCAGGACCTTCATTATGAA-3'.

For amplification from tissue: *RPL28* forward GGTGCAGGTTAGGTGGACTG-3', *RPL28* Rv CGGATCATGTGTCTGATGCTG-3'.

For amplicon-based next-generation sequencing: *SEPT2* forward p1 5'-TGTAACGACGCGCCAGTCTCTCTCTTTCAGCACCCA-3', *SEPT2* reverse p1 5'-CAGGAAACAGCTATGACCTGGCTCAGGGTGACACAATAC-3', *SEPT2* forward p2 5'-TGTAACGACGCGCCAGTGCCACCTTGGTGATTCTCTC-3', *SEPT2* reverse p2 5'-CAGGAAACAGCTATGACACACAATACAGAGAAAGGGGCA-3'; *EML1* forward p1 5'-TGTAACGACGCGCCAGT

GACAGACCGCATTTGCTTCAC-3', *EML1* reverse p1 5'-CAGGAAACAGCTATGACCCCTTTGGTAGGTCCTTTCCTGT-3', *EML1* forward p2 5'-TGTAACGACGCGCCAGTTCAGATGCAAGAAGACGACA-3', *EML1* reverse p2 5'-CAGGAAACAGCTATGACCGTTCTGTGTCCTCAAACGC-3'; *PDCD10* forward p1 5'-TGTAACGACGCGCCAGTAGCAGACGAATAAAGCAGGA-3', *PDCD10* reverse p1 5'-CAGGAAACAGCTATGACCTGAAGCTGAGACCACATCCAT-3', *PDCD10* forward p2 5'-TGTAACGACGCGCCAGTGAGGAATTAAGAATTGCAGAGTT-3', *PDCD10* reverse p2 5'-CAGGAAACAGCTATGACGAGAATGAAGCTGAGACCACATC-3'.

For loss of heterozygosity. *AFMA357YH1* forward (HEX) 5'-CATAGGCCAGCAAGCTCAA-3', *AFMA357YH1* reverse 5'-ACGTGTCTTCTTCGTACCC-3'; *AFM182YA5* forward (HEX) 5'-TGACACGTGAACAGACTAAGCA-3', *AFM182YA5* reverse 5'-CAATACGGGAGAGCCAGTTGT-3'; *D19S1142* forward (HEX) 5'-AGGACGGAAAGGCAGAGAAA, *D19S1142* reverse 5'-CCCTCTGATCCTCTTTGCT-3'; *RH56150* forward (HEX) 5'-AGAAAGGTCGAGGGACTCA-3', *RH56150* reverse 5'-GCCAGTTAATTTGGAAGCAGGT-3'.

Statistics. As indicated in the corresponding figure legends, statistical significance was calculated by GraphPad Prism 6.0c software using the unpaired, two-tailed Student's *t*-test and *P* values <0.05 were considered significant, unless otherwise indicated.

eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation

Amy S. Y. Lee^{1,2,†}, Philip J. Kranzusch^{1,2,3,†}, Jennifer A. Doudna^{1,2,3,4,5,6} & Jamie H. D. Cate^{1,2,4,5}

Eukaryotic mRNAs contain a 5' cap structure that is crucial for recruitment of the translation machinery and initiation of protein synthesis. mRNA recognition is thought to require direct interactions between eukaryotic initiation factor 4E (eIF4E) and the mRNA cap. However, translation of numerous capped mRNAs remains robust during cellular stress, early development, and cell cycle progression¹ despite inactivation of eIF4E. Here we describe a cap-dependent pathway of translation initiation in human cells that relies on a previously unknown cap-binding activity of eIF3d, a subunit of the 800-kilodalton eIF3 complex. A 1.4 Å crystal structure of the eIF3d cap-binding domain reveals unexpected homology to endonucleases involved in RNA turnover, and allows modelling of cap recognition by eIF3d. eIF3d makes specific contacts with the cap, as exemplified by cap analogue competition, and these interactions are essential for assembly of translation initiation complexes on eIF3-specialized mRNAs² such as the cell proliferation regulator *c-Jun* (also known as *JUN*). The *c-Jun* mRNA further encodes an inhibitory RNA element that blocks eIF4E recruitment, thus enforcing alternative cap recognition by eIF3d. Our results reveal a mechanism of cap-dependent translation that is independent of eIF4E, and illustrate how modular RNA elements work together to direct specialized forms of translation initiation.

The rate-limiting step of translation initiation is the recognition of the 5' cap structure by eIF4E^{3,4}. eIF4E activity is highly regulated by extracellular stimuli, predominantly through steric hindrance by eIF4E-binding proteins (4E-BPs)^{5,6}. The translational efficiencies of mRNAs range in sensitivity to 4E-BP inhibition^{7–9}, and these differences have conventionally been addressed by categorizing translation into cap-dependent versus cap-independent pathways¹⁰. However, the mechanisms underlying mRNA sensitivity to active eIF4E levels remain enigmatic as all cellular mRNAs maintain the same 5' cap structure¹¹.

Recently, we discovered a new translation pathway driven by RNA interactions with eIF3 that is used by a subset of cell proliferation mRNAs, with the prototype member being the mRNA encoding the early response transcription factor *c-Jun*². eIF3-specialized translation is cap-dependent and requires recruitment of eIF3 to an internal stem-loop structure in the 5' untranslated region (UTR). However, the translational efficiency of a subset of these mRNAs is unaffected by eIF4E inactivation^{7–9}, suggesting that cap recognition may proceed by a non-canonical mechanism (Supplementary Table 1).

To understand how cap recognition occurs during eIF3-specialized translation, we examined whether *c-Jun* mRNA uses the canonical eIF4F cap-binding complex during initiation. We programmed *in vitro* translation extracts from human 293T cells with capped and polyadenylated *c-Jun* mRNA, and isolated the 48S complex to assess the presence of the eIF4F factors (eIF4G1, eIF4A1 and eIF4E) (Fig. 1a, b). Unexpectedly, although *c-Jun* mRNA translation initiation complexes contain eIF3 and the small ribosomal subunit, they are depleted of

all eIF4F components. By contrast, eIF4F is readily detectable in 48S initiation complexes formed on a canonical eIF4E-dependent mRNA, *ACTB*¹² (Fig. 1b). In agreement with the absence of eIF4F, *c-Jun* levels are unaffected by cell treatment with the mTOR inhibitor INK128 (ref. 7), which inactivates eIF4E, or with eIF4A inhibitors¹³ (Extended Data Fig. 1). These results indicate that *c-Jun* mRNA translation occurs independently of eIF4F and that the process of eIF3-specialized translation is fundamentally distinct at the initial stage of 5' cap recognition.

eIF3-specialized translation requires recognition of an internal RNA stem-loop for efficient translation². Therefore, we asked whether eIF3

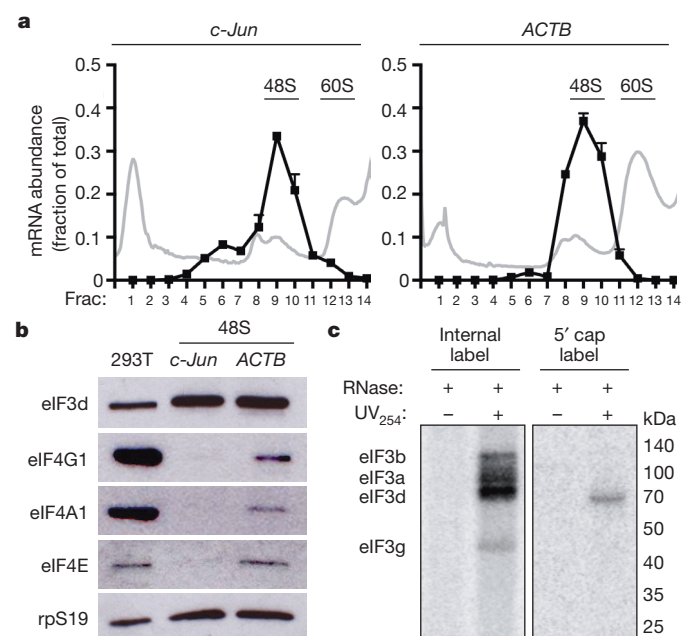


Figure 1 | 5' end recognition of *c-Jun* mRNA is eIF4F-independent.

a, Distribution of *c-Jun* or *ACTB* mRNA-containing initiation complexes in programmed 293T cell *in vitro* translation extracts. The mRNA abundance (black line) is expressed as the fraction of total recovered transcripts. The results are given as the mean \pm s.d. of a representative quantitative RT-PCR experiment performed in duplicate. The polysome profile (grey line) is plotted as relative absorbance at 254 nm versus elution fractions. **b**, Western blot analysis of initiation factors in 48S translation complexes formed on *c-Jun* and *ACTB* mRNAs. 293T, total protein from 293T *in vitro* translation extracts. rpS19, ribosomal protein S19. For gel source data, see Supplementary Fig. 1. **c**, Phosphorimage of SDS-PAGE gel resolving RNase-protected ³²P-internal or ³²P-cap-labelled *c-Jun* 5' UTR RNA crosslinked to eIF3 subunits. Recombinant eIF3a migrates at ~100 kDa owing to a C-terminal truncation²⁶. The results of a–c are representative of three independent experiments.

¹Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. ²Center for RNA Systems Biology, University of California, Berkeley, Berkeley, California 94720, USA. ³Howard Hughes Medical Institute (HHMI), University of California, Berkeley, Berkeley, California 94720, USA. ⁴Department of Chemistry, University of California, Berkeley, Berkeley, California 94720, USA. ⁵Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁶Innovative Genomics Initiative, University of California, Berkeley, California 94720, USA. [†]Present addresses: Department of Biology, Brandeis University, Waltham, Massachusetts 02454, USA (A.S.Y.L.); Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA (P.J.K.); Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA (P.J.K.).

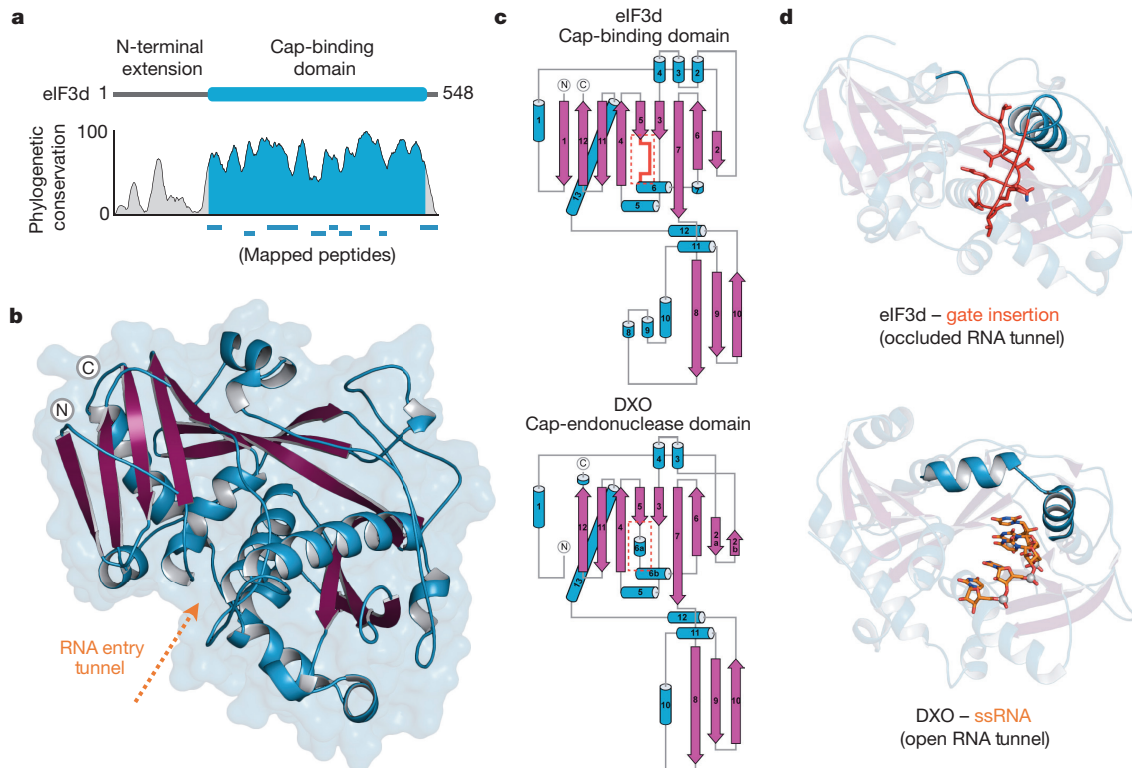


Figure 2 | Structure of eIF3d reveals a conserved cap-binding domain.
a, Cartoon schematic and phylogenetic conservation of eIF3d amino acid sequence according to physiochemical property similarity. Peptides in the cap-binding domain as identified by limited proteolysis are mapped below.
b, Structure of the eIF3d cap-binding domain. α -helices are coloured

might also be involved in 5' cap recognition. In agreement with the previously demonstrated RNA-binding capability of eIF3, the four eIF3 RNA-binding subunits, eIF3a, eIF3b, eIF3d and eIF3g, provide RNase protection to internally ^{32}P -labelled *c-Jun* 5' UTR RNA after UV₂₅₄-induced crosslinking² (Fig. 1c). By contrast, when the ^{32}P label is placed in the 5' cap of *c-Jun* mRNA, RNase protection is observed with a single subunit of eIF3, corresponding to eIF3d (Fig. 1c, Extended Data Fig. 2a). We confirmed subunit identity by limited proteolysis and mass spectrometry, and defined a C-terminal region of eIF3d that is responsible for protection of the 5' mRNA terminus (Extended Data Fig. 2). The mapped C-terminal region of eIF3d is broadly conserved throughout plant, fungal and animal phylogeny (Fig. 2a, Extended Data Fig. 3), suggesting the apparent 5' end recognition activity of eIF3d is an evolutionarily preserved function of the eIF3 complex.

To understand how eIF3d recognizes the 5' RNA terminus, we determined a 1.4 Å crystal structure of the conserved C-terminal domain of eIF3d from *Nasonia vitripennis* (65% identical, 84% similar to human eIF3d) using sulfur anomalous dispersion for phase determination (Extended Data Table 1, Extended Data Fig. 3). The structure of eIF3d reveals a complex fold that forms a cup-shaped architecture with a positively charged central tunnel that is negatively charged at its base (Fig. 2b). Remarkably, despite no significant sequence homology, the structural topology of eIF3d is nearly identical to the DXO proteins, a recently described family of 5' cap-endonucleases involved in RNA quality control^{14–16} (Fig. 2c, Extended Data Fig. 4). In contrast to DXO, eIF3d contains a unique insertion of ~15 highly conserved amino acids between strand β 5 and helix α 6. The eIF3d-specific insertion folds down along the front face of the domain, making loosely packed charged interactions that close off the RNA binding tunnel (Extended Data Fig. 5). We term this insertion an 'RNA gate', as the sequence clashes with the path of single-stranded RNA (ssRNA) bound to DXO¹⁵ and must undergo a conformational change for eIF3d to

become competent for RNA recognition (Fig. 2d). We determined the structure of eIF3d in two additional crystal forms, and confirmed the RNA gate exhibits a closed conformation regardless of crystal packing (Extended Data Fig. 6). As eIF3d does not bind all capped RNAs^{17,18}, we postulate that the RNA gate regulates cap recognition to prevent promiscuous mRNA binding before assembly of eIF3d into the full eIF3 complex. We tested this model using *c-Jun* mRNA, and verified that eIF3d cap-recognition only occurs in the context of a full eIF3 complex and requires previous eIF3-sequence-specific RNA interactions with the eIF3-recruitment stem-loop (Extended Data Fig. 7). Allosteric communication between eIF3 subunits during initial RNA recruitment likely facilitates eIF3d RNA gate opening to allow 5' end recognition. The structure of eIF3d therefore reveals a new cap-binding protein and explains the ability of the eIF3 complex to protect the 5' end of mRNA (Fig. 1c).

To validate the structural finding that eIF3d is a cap-binding protein, we examined the ability of eIF3 to bind the *c-Jun* mRNA 5' cap in the presence of competitor ligands. eIF3d cap recognition is sensitive to m⁷GDP competition but resistant to GDP, indicating that, analogous to eIF4E⁴, eIF3d specifically interacts with the 5' cap and requires a mature methylated cap structure for recognition (Fig. 3a). Using the DXO–RNA structure as a template¹⁵, we modelled a capped ssRNA along the basic binding groove shared between eIF3d and DXO and identified two conserved helices (α 5 and α 11) likely to be involved in cap recognition (Fig. 3b). We purified recombinant eIF3 containing helix α 5- or α 11-mutated eIF3d and demonstrated that both mutants have markedly reduced ability to crosslink to the *c-Jun* mRNA cap (Fig. 3c). eIF3d-mutated complexes retain wild-type-levels of RNA-binding, indicating that these residues specifically coordinate 5' mRNA cap recognition (Extended Data Fig. 8). We next introduced haemagglutinin (HA) epitope-tagged wild-type or mutant eIF3d into 293T cells, and measured the assembly of 48S initiation complexes on *c-Jun*

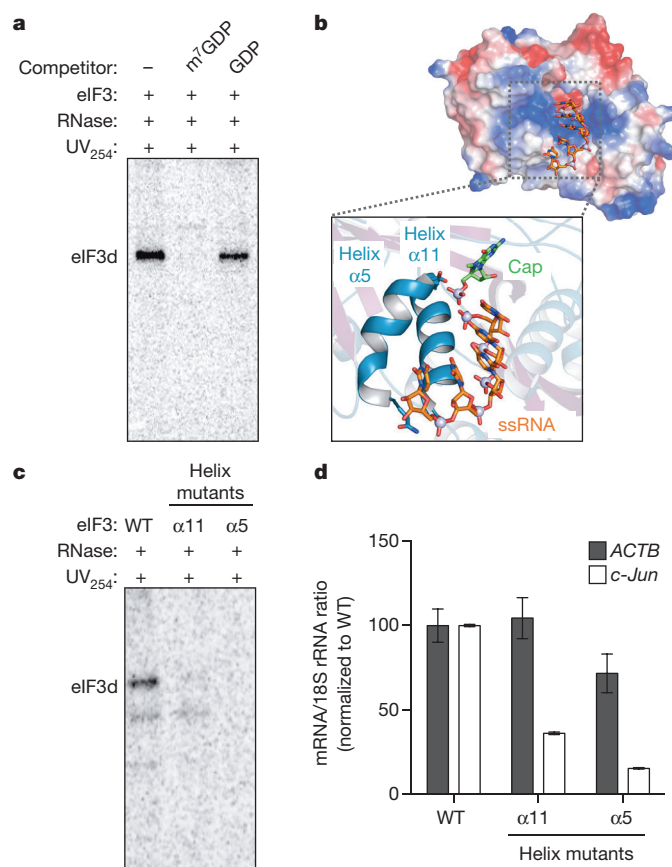


Figure 3 | eIF3d cap-binding activity is required for efficient 48S initiation complex formation on specific mRNAs. **a**, Phosphorimage of SDS-PAGE gel resolving RNase-protected 32 P-cap-labelled *c-Jun* 5' UTR RNA crosslinked to eIF3 in the presence of competitor ligands (m⁷GDP, GDP). **b**, Electrostatic surface view of the eIF3d cap-binding domain coloured by charge, with a zoomed view of ssRNA and cap analogue modelled according to their positions bound to DXO¹⁵. Positive charge is coloured blue, negative charge is in red, and the RNA gate is removed for clarity. **c**, Phosphorimage of SDS-PAGE gel resolving RNase-protected 32 P-cap-labelled *c-Jun* 5' UTR RNA crosslinked to wild-type (WT) or helix α 5- or helix α 11-mutant eIF3. Helix α 5-mutant eIF3d: D249Q/V262I/Y263A; helix α 11-mutant eIF3d: T317E/N320E/H321A. **d**, Incorporation of *c-Jun* and *ACTB* mRNA into initiation complexes by wild-type, helix α 5-, or helix α 11-mutant eIF3d as measured by quantitative RT-PCR. The mRNA-ribosome association is expressed as the ratio of the quantity of mRNA transcripts to 18S rRNA and normalized to the wild-type sample. The results are representative of three independent experiments and given as the mean \pm s.d. from a representative quantitative RT-PCR experiment performed in duplicate.

mRNA by quantitative RT-PCR^{19,20}. Mutations to the predicted eIF3d cap-binding surface inhibit *c-Jun* mRNA incorporation into translation complexes, while the control *ACTB* mRNA is unaffected (Fig. 3d, Extended Data Fig. 8). These results demonstrate that cap binding by eIF3d is required for efficient initiation complex formation during eIF3-specialized translation.

eIF3d recognition of the 5' cap structure provides an alternative cap-dependent translation mechanism from canonical eIF4F cap recognition. Perplexingly, when the RNA stem-loop element that recruits eIF3 to the *c-Jun* mRNA is deleted, translation is inhibited even though the mRNA contains a 5' cap². We proposed that an RNA element within the *c-Jun* mRNA blocks recruitment of the eIF4F complex. In support, the 5' cap of *c-Jun* mRNA crosslinks less efficiently to purified eIF4E than that of the *ACTB* mRNA (Extended Data Fig. 9). To identify the eIF4F inhibitory element, we constructed luciferase reporters to test deletions in the *c-Jun* 5' UTR (Fig. 4a). Deletion of

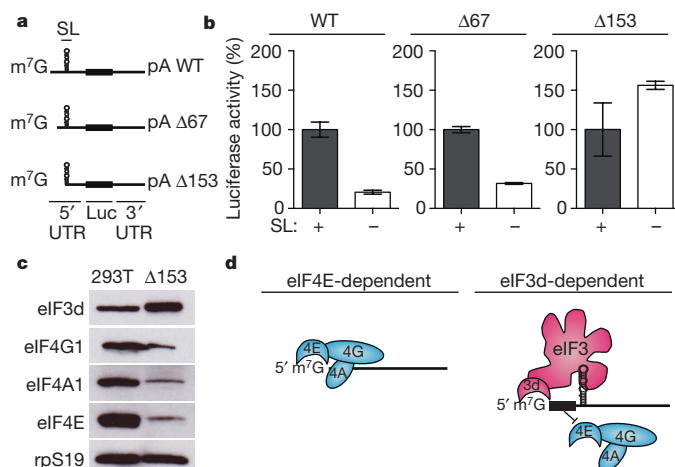


Figure 4 | An RNA element inhibits eIF4F recruitment and directs mRNAs to use an eIF3-specialized translation pathway. **a**, Schematic of *c-Jun* 5' UTR truncation-luciferase (Luc) reporter mRNAs. SL, stem-loop. **b**, Luciferase activity from *in vitro* translation of mRNAs containing truncations of the *c-Jun* 5' UTR, with or without the internal eIF3-recruitment stem-loop sequence. The results are given as the mean \pm s.d. of three independent experiments, each performed in triplicate. **c**, Western blot analysis of initiation factors in 48S translation initiation complexes formed on *c-Jun* mRNA with a 5' 153-nucleotide truncation. 293T, total protein from 293T *in vitro* translation extracts. The result is representative of three independent experiments. For gel source data, see Supplementary Fig. 1. **d**, Model for eIF3d-directed cap-dependent mRNA translation. An eIF4F-inhibitory RNA element ensures that mRNA translation occurs through an eIF3-specialized pathway.

the 5' 153 nucleotides, but not the initial 67 nucleotides, was sufficient to allow *c-Jun* mRNA translation to occur independently of the eIF3-recruitment stem-loop, suggesting that canonical cap dependent translation is no longer blocked (Fig. 4b). We confirmed by western blot analysis of the 48S initiation complex formed on *c-Jun* mRNA lacking the 5' 153 nucleotides that the eIF4F components are now present (Fig. 4c).

Together, we put forth a model of a previously undiscovered cap-dependent translation initiation pathway controlled by eIF3d recognition of the 5' mRNA cap (Fig. 4d). We postulate that encoding more than one mechanism of cap-dependent translation allows cells to control protein synthesis specifically in cellular environments in which eIF4E is inactivated. In support, *c-Jun* mRNA translation is resistant to treatment of cells with chemicals that activate the 4E-BPs^{7,8,21} (Extended Data Fig. 1). As modulation of eIF4E cap-binding activity allows cells to incorporate extracellular stimuli into altered translation outputs²², it will be important to discover whether eIF3d activity is analogously regulated. Furthermore, our data indicates that the *c-Jun* mRNA encodes an additional *cis*-acting RNA element that blocks eIF4F to ensure translation can only occur through an eIF3-specialized pathway. RNA elements that block eIF4F recruitment may be a common theme to direct mRNAs into specific translation pathways to ensure controlled protein expression. For example, a subset of homeobox mRNAs contain an RNA element that blocks cap-dependent translation to ensure usage of an internal ribosome entry site and to allow for correct homeobox expression during embryonic development²³. Several eIF3-specialized mRNAs encode proteins involved in the control of cell proliferation, suggesting that their translation may also require enhanced regulation^{2,24}.

While considerable advances have been made in the structural understanding of eIF3 bound to the ribosome, direct localization of eIF3d in a 48S complex remains unclear²⁵. Thus, understanding how eIF3d functions and assembles within the full translation initiation complex will have important mechanistic implications in how cap recognition links to mRNA ribosomal recruitment. Our discovery of

eIF3d as a cap-binding protein now reveals a new translation pathway independent of eIF4E, and adds another layer of cap-dependent translation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 31 December 2015; accepted 21 June 2016.

Published online 27 July 2016.

- Gingras, A. C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963 (1999).
- Lee, A. S., Kranzusch, P. J. & Cate, J. H. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**, 111–114 (2015).
- Sonenberg, N. eIF4E, the mRNA cap-binding protein: from basic discovery to translational research. *Biochem. Cell Biol.* **86**, 178–183 (2008).
- Sonenberg, N., Morgan, M. A., Merrick, W. C. & Shatkin, A. J. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc. Natl Acad. Sci. USA* **75**, 4843–4847 (1978).
- Pause, A. *et al.* Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* **371**, 762–767 (1994).
- Gingras, A. C. *et al.* Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.* **13**, 1422–1437 (1999).
- Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).
- Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
- Gandin, V. *et al.* nanoCAGE reveals 5' UTR features that define specific modes of translation of functionally related MTOR-sensitive mRNAs. *Genome Res.* **26**, 636–648 (2016).
- Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
- Shatkin, A. J. Capping of eucaryotic mRNAs. *Cell* **9**, 645–653 (1976).
- Andreev, D. E. *et al.* Differential contribution of the m7G-cap to the 5' end-dependent translation initiation of mammalian mRNAs. *Nucleic Acids Res.* **37**, 6135–6147 (2009).
- Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
- Chang, J. H. *et al.* Dxo1 is a new type of eukaryotic enzyme with both decapping and 5'-3' exoribonuclease activity. *Nat. Struct. Mol. Biol.* **19**, 1011–1017 (2012).
- Jiao, X., Chang, J. H., Kilic, T., Tong, L. & Kiledjian, M. A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol. Cell* **50**, 104–115 (2013).
- Xiang, S. *et al.* Structure and function of the 5'→3' exoribonuclease Rat1 and its activating partner Rai1. *Nature* **458**, 784–788 (2009).
- Sonenberg, N., Morgan, M. A., Testa, D., Colonna, R. J. & Shatkin, A. J. Interaction of a limited set of proteins with different mRNAs and protection of 5'-caps against pyrophosphatase digestion in initiation complexes. *Nucleic Acids Res.* **7**, 15–29 (1979).
- Lindqvist, L., Imataka, H. & Pelletier, J. Cap-dependent eukaryotic initiation factor-mRNA interactions probed by cross-linking. *RNA* **14**, 960–969 (2008).
- Locker, N., Easton, L. E. & Lukavsky, P. J. Affinity purification of eukaryotic 48S initiation complexes. *RNA* **12**, 683–690 (2006).
- Zhang, L., Pan, X. & Hershey, J. W. Individual overexpression of five subunits of human translation initiation factor eIF3 promotes malignant transformation of immortal fibroblast cells. *J. Biol. Chem.* **282**, 5790–5800 (2007).
- Blau, L. *et al.* Aberrant expression of c-Jun in glioblastoma by internal ribosome entry site (IRES)-mediated translational activation. *Proc. Natl Acad. Sci. USA* **109**, E2875–E2884 (2012).
- Hay, N. & Sonenberg, N. Upstream and downstream of mTOR. *Genes Dev.* **18**, 1926–1945 (2004).
- Xue, S. *et al.* RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature* **517**, 33–38 (2015).
- Wisdom, R., Johnson, R. S. & Moore, C. c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J.* **18**, 188–197 (1999).
- des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525**, 491–495 (2015).
- Sun, C. *et al.* Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc. Natl Acad. Sci. USA* **108**, 20473–20478 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank J. Berger and K. Chat for discussions. X-ray data were collected at Beamline 8.3.1 of the Lawrence Berkeley National Laboratory Advanced Light Source (ALS), supported in part by the UC Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599 and the Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. The authors are grateful to J. Holton, G. Meigs (ALS), and T. Doukov (SSRL) for help with S-SAD data collection. This work used the Vincent J. Proteomics/Mass Spectrometry Laboratory at UC Berkeley, supported in part by NIH S10 Instrumentation Grant S10RR025622. This work was funded by the NIGMS Center for RNA Systems Biology (P50-GM201706). J.A.D. is an HHMI Investigator. A.S.Y.L. is supported as an American Cancer Society Postdoctoral Fellow (PF-14-108-01-RMC) and P.J.K. is supported as an HHMI Fellow of the Life Sciences Research Foundation.

Author Contributions Project and experiments were conceived and designed by A.S.Y.L. with consultation with J.H.D.C. Protein purification, biochemistry, and cell biology experiments and analyses were performed by A.S.Y.L. Crystallography and structure determination was performed by P.J.K. The manuscript was written by A.S.Y.L., P.J.K. and J.H.D.C. with editing by J.A.D.

Author Information Atomic coordinates and structure factors for the reported crystal structures are deposited at the Protein Data Bank (PDB) under accession codes 5K4B (crystal form 1), 5K4C (crystal form 2), and 5K4D (crystal form 3). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.D.C. (jjcate@lbl.gov).

Reviewer Information *Nature* thanks I. Topisirovic and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Cells and transfections. Human 293T cells were maintained in DMEM (Invitrogen) supplemented with 10% FBS (Seradigm). The cells were obtained from the University of California, Berkeley, Cell Culture Facility, which authenticates cells by STR profiling and tests for mycoplasma contamination. Plasmid transfections were performed using Lipofectamine 2000 (Invitrogen), following the manufacturer's protocol, and polysome or immunoprecipitation analyses were performed at 48 h after transfection. For INK128 (Cayman Chemical) cell treatment, 293T cells were incubated with the indicated concentration of INK128 for ~14–16 h before cell lysis.

Plasmids. To generate the eIF3d expression plasmids, eIF3d was amplified from human cDNA and inserted into pcDNA5/FRT. A 39-nucleotide linker followed by the HA epitope tag (YPYDVPDYA) was subsequently inserted before the eIF3d stop codon. The wild-type *c-Jun* 5' UTR luciferase reporter plasmid was previously described². To generate the *c-Jun* *in vitro* transcription template, the 5' UTR, ORF and 3' UTR were separately amplified from human cDNA and stitched together downstream of a T7 promoter by Gibson cloning into pcDNA4. The *ACTB* *in vitro* transcription template was constructed by addition of a T7 promoter during amplification of the full mRNA from human cDNA and inserted into pcDNA4.

Western blot. Western blot analyses were performed using the following antibodies: anti-eIF3d (Bethyl A301-758A), anti-eIF4A1 (Cell Signaling 2490), anti-eIF4G1 (Cell Signaling 2858), anti-rpS19 (Bethyl A304-002A), anti-eIF4E (Bethyl A301-154A), anti-rpLp0 (Bethyl A302-882A), anti-HA epitope tag (Pierce 26183), anti-*c-Jun* (Cell Signaling 9165), anti-Hsp90 (BD Biosciences 610418), and anti-4E-BP1 (Cell Signaling 9644).

***In vitro* RNA transcription and labelling.** Unlabelled RNAs were *in vitro* transcribed, polyadenylated, and capped as previously described². For internal radiolabelling of RNAs, *in vitro* transcription was performed in the presence of 0.1 μ M [α -³²P]ATP, then the RNA was subsequently capped with vaccinia virus enzymes (NEB). For radiolabelling of the 5' cap, *in vitro* transcribed RNAs were capped with vaccinia virus enzymes and [α -³²P]GTP. RNAs were purified by phenol–chloroform extraction and ethanol precipitation.

***In vitro* translation.** *In vitro* translation extracts were made from human 293T cells as previously described²¹. Lysates were nuclease-treated with 18 gel U μ l⁻¹ micrococcal nuclease (NEB) in the presence of 0.7 mM CaCl₂ for 10 min at 25°C, and the digestion was stopped by addition of 2.24 mM EGTA. Each translation reaction contained 50% *in vitro* translation lysate and buffer to make the final reaction with 0.84 mM ATP, 0.21 mM GTP, 21 mM creatine phosphate (Roche), 45 U ml⁻¹ creatine phosphokinase (Roche), 10 mM HEPES-KOH, pH 7.6, 2 mM DTT, 8 mM amino acids (Promega), 255 mM spermidine, 1 U ml⁻¹ murine RNase inhibitor (NEB), and mRNA-specific concentrations of Mg(OAc)₂ and KOAc. The optimal magnesium and potassium levels to add were determined to be 1.5 mM Mg(OAc)₂ and 150 mM KOAc for *c-Jun* mRNA, and 1 mM and 150 mM KOAc for *ACTB* mRNA. For luciferase assays, translation reactions were incubated for 1 h at 30°C, then luciferase activity was assayed.

48S initiation complex purification. For 48S initiation complex purification from *in vitro* translation reactions, 180 μ l reactions were incubated in the presence of GMP-PNP for 20 min at 30°C and centrifuged for 6 min at 12,000g at 4°C. Lysates were purified by size-exclusion chromatography through a 1 ml column packed with Sephacryl S-400 gel filtration resin (GE Healthcare) and the elutant was centrifuged through a 10–25% (w/v) sucrose gradient by centrifugation for 3.5 h at 38,000 r.p.m. at 4°C in a Beckman SW41 Ti rotor²⁷. Fractions were collected from the top of the gradient using a peristaltic pump with a Brandel tube piercer. From the appropriate fractions, RNA was purified by phenol–chloroform extraction and ethanol precipitation and protein was precipitated with trichloroacetic acid.

For affinity purification of HA epitope-tagged eIF3d-associated 48S initiation complexes from cells, three 10 cm plates of transfected 293T cells were treated with 100 μ g ml⁻¹ cycloheximide for 5 min. Cells were washed with ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 100 mM Na₂HPO₄, 2 mM KH₂PO₄) with 100 μ g ml⁻¹ cycloheximide and collected in lysis buffer (20 mM HEPES-KOH pH 7.4, 150 mM KOAc, 2.5 mM Mg(OAc)₂, 1 mM DTT, 100 μ g ml⁻¹ cycloheximide, 1% (v/v) Triton X-100). Lysates were centrifuged for 6 min at 12,000g at 4°C and purified by S-400 size-exclusion chromatography. 80 μ l of anti-HA antibody-conjugated agarose beads (Sigma) was added to the elutants, tumbled for 1.5 h at 4°C, and beads were washed three times with lysis buffer without Triton X-100. Bound complexes were eluted twice with 100 μ g ml⁻¹ HA peptide and elutants were centrifuged and analysed the same as for the *in vitro* purification reactions.

Quantitative real-time PCR. cDNA was reverse-transcribed from RNA using random hexamers and Superscript III (Invitrogen), following the manufacturer's protocol. Real-time PCR was performed using DyNamo HS Sybr Green (ThermoFisher), with a 20 μ l reaction volume containing 2 μ l cDNA and 0.5 μ M of each primer. The following oligonucleotides were used: 18S rRNA forward, 5'-GGCCCTGTAATTGGAATGAGTC-3', 18S rRNA reverse, 5'-CCAAGA

TCCAACACTACGAGCTT-3'; *ACTB* forward, 5'-CTCTTCCAGCCTTCCTTCCT-3', *ACTB*-reverse, 5'-AGCACTGTGTGGCGTACAG-3'; *c-Jun* forward, 5'-TGACTGCAAAGATGGAACG-3', *c-Jun* reverse, 5'-CAGGGTCATGCTCTGTTTCA-3'.

eIF3-RNA crosslinking and gel shift. Recombinant eIF3 was expressed and purified from *Escherichia coli* as previously described²⁶. For each crosslinking reaction, 1 μ l water, 1 μ l 125 nM labelled RNA, 1 μ l 10 μ g ml⁻¹ heparan sulfate (Sigma), 1 μ l 5 \times binding buffer (125 mM Tris-HCl, pH 7.5, 25 mM Mg(OAc)₂, 350 mM KCl, 0.5 mM CaCl₂, 0.5 mg ml⁻¹ BSA, 10 mM TCEP), and 1 μ l 1.5 μ M purified eIF3 were added, in the listed order, and incubated for 30 min at 25°C. For competition experiments, the water was substituted with 1 μ l of 1 mM m⁷GDP/Mg²⁺ or GDP/Mg²⁺. UV₂₅₄-induced crosslinking was performed using a short-wave UV lamp placed ~4 cm above the samples on ice for 10 min. After treatment with RNase for 10 min at 37°C, proteins were separated by 12% SDS-PAGE, the gel was dried, and imaged using a phosphorimager. For digestion of internal labelled RNA, 2.5 U benzonase (Novagen) and 250 U RNase T1 (ThermoFisher) were used; for digestion of cap labelled RNA, 4 U RNase R (Epicentre) and 1 U RiboShredder (Epicentre) were used. For eIF3d subunit identification, after RNase treatment, samples were denatured and immunoprecipitation was performed as previously described². For limited proteolysis, after RNase treatment, the reactions were treated with 2 or 20 μ g ml⁻¹ sequencing grade trypsin (Promega) for 30 min at 25°C, before gel electrophoresis. Mass spectrometry samples were prepared as previously described². Gel-shift assays were performed as previously described, using 50 nM labelled *c-Jun* stem-loop RNA and 300 nM purified eIF3 (ref. 2).

Recombinant eIF3d protein purification. Candidate eIF3d cap-binding domain fragments were amplified by PCR and cloned into a modified pET vector to express an N-terminal 6 \times His (KSSHHHHHHGSS)-MBP-TEV fusion protein as previously described²⁸. Extensive expression trials were conducted to determine optimal N- and C-terminal domain boundaries and identified a minimal stable human cap-binding domain S161–F527. Recombinant protein was expressed in BL21-RIL DE3 *E. coli* cells co-transformed with a pRARE2 tRNA plasmid (Agilent). *E. coli* was grown in 2 \times YT media at 37°C to an OD₆₀₀ of ~0.5, cooled at 4°C for 15 min, induced with addition of 0.5 mM IPTG and then incubated with shaking for ~20 h at 16°C. Pelleted cells were washed with PBS and then lysed by sonication in lysis buffer (20 mM HEPES-KOH pH 7.5, 400 mM NaCl, 10% glycerol, 30 mM imidazole, 1 mM TCEP) in the presence of EDTA-free Complete Protease Inhibitor (Roche). Following centrifugation for 30 min at 23,000g and 4°C, clarified lysate was incubated with Ni-NTA agarose resin (QIAGEN) for 1 h at 4°C with gentle rocking. Resin was washed with lysis buffer supplemented to 1 M NaCl and eluted by gravity-flow chromatography at 4°C with lysis buffer supplemented to 300 mM imidazole. The eluted fraction was diluted to ~50 mM imidazole and 5% glycerol, concentrated to ~50 mg ml⁻¹ and incubated with Tobacco Etch Virus protease for ~12 h at 4°C to remove the MBP tag. Recombinant eIF3d was isolated from free MBP by diluting with gel-filtration buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM TCEP) and passing over a 5 ml Ni-NTA column (QIAGEN) connected in line with a 5 ml MBP-Trap column (GE Life Sciences) before additional purification by size-exclusion chromatography on a Superdex 75 16/60 column. Final purified eIF3d was concentrated to ~20–50 mg ml⁻¹, used immediately for crystallography, or flash frozen in liquid nitrogen for storage at –80°C.

Crystallization and structure determination. Initial crystals of human eIF3d were grown at 18°C by hanging drop vapour diffusion, but diffracted poorly. Analogous eIF3d cap-binding domain sequences were cloned from a panel of highly homologous animal sequences, with the equivalent domain from the parasitic wasp *N. vitripennis* (S172–F537) producing the best crystals. Optimized *N. vitripennis* eIF3d crystals were grown in 2 μ l hanging drops set at a 1:1 ratio over 300 μ l of reservoir liquid: 200 mM (NH₄)₂SO₄, 100 mM Bis-Tris 6.5, 23–27% PEG-3350 (crystal form 1), 1.6–1.8 M ammonium citrate, pH 7.0 (crystal form 2), or 200 mM NaCl, 100 mM Tris 8.5, 25% PEG-3350 (crystal form 3). eIF3d crystals (crystal forms 1 and 2) were cryoprotected by covering the drop with a layer of saturated paratone-N or NVH oil (Hampton) and crystals were transferred into the oil emersion and cleaned using a Kozak cat whisker as previously described²⁹, or cryoprotected by transferring to a reservoir solution supplemented with 20% ethylene glycol (crystal form 3). Crystals were harvested with a nylon loop and then flash-frozen in liquid nitrogen. X-ray diffraction data were collected under cryogenic conditions at the Lawrence Berkeley National Laboratory Advanced Light Source (beamline 8.3.1).

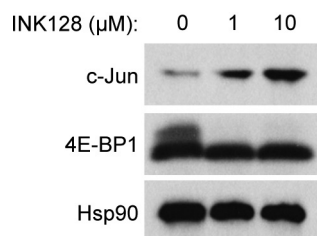
Data were processed with XDS and AIMLESS³⁰ using the SSRL *autoxds* script (A. Gonzalez, Stanford SSRL). eIF3d crystals belonged to the orthorhombic space group P2₁ 2₁ 2₁, and contained either two copies per asymmetric unit (crystal form 1) or one copy (crystal form 2), or the space group P2₁ and contained two copies per asymmetric unit (crystal form 3). Experimental phase information was collected from a native crystal using sulfur single-wavelength anomalous dispersion. Data were collected at a minimal accessible wavelength (~7.235 eV) and iterative data

sets were completed and merged from independent portions of an exceptionally large eIF3d crystal. After $\sim 90\times$ multiplicity, anomalous signal was detected to $\sim 2.4\text{ \AA}$, and a clear phase solution was obtained at $\sim 120\times$ multiplicity. 35 sites were identified with HySS in PHENIX³¹ corresponding to 32 sulfur atoms in eIF3d and 3 chloride positions. Phases were extended to the native eIF3d data set processed to $\sim 1.40\text{ \AA}$ using SOLVE/RESOLVE³², and model building was completed in Coot³³ before refinement with PHENIX. X-ray data for refinement were extended according to an I/σ resolution cut-off of ~ 1.5 , CC* correlation and R_{pim} parameters, and visual inspection of the resulting map³⁴. A completed eIF3d cap-binding domain from crystal form 1 was used as a search model to determine phases for crystal form 2 and 3 using molecular replacement. Final structures were refined to stereochemistry statistics for Ramachandran plot (favoured/allowed), rotamer outliers, and MolProbity score as follows: crystal form 1, 96.8%/3.2%, 0.2% and 1.40; crystal form 2, 97.3%/2.7%, 0% and 1.29; crystal form 3 97.4%/2.6%, 0.9% and 1.26.

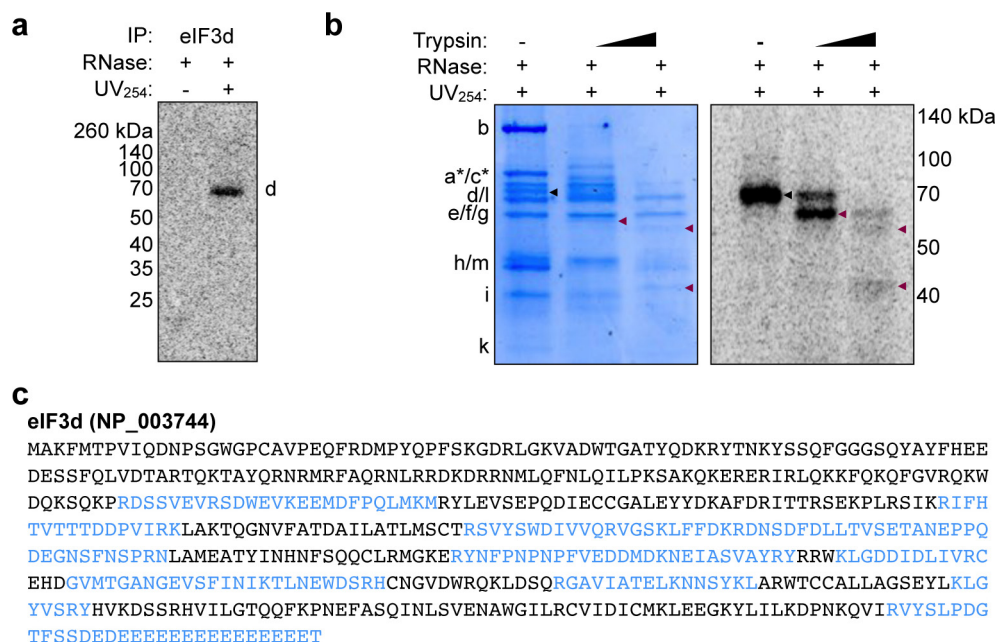
Recombinant eIF4E protein purification and RNA crosslinking. Full-length human eIF4E was cloned and expressed using the same protocol as for eIF3d. eIF4E–RNA crosslinking was performed as described for eIF3–RNA crosslinking,

but using 25 nM RNA with normalized counts per million and the indicated concentration of eIF4E. RNase treatment was performed using 4 U RNase R and 250 U RNase T1.

27. Lee, A. S., Burdeinick-Kerr, R. & Whelan, S. P. A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs. *Proc. Natl. Acad. Sci. USA* **110**, 324–329 (2013).
28. Kranzusch, P. J. *et al.* Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* **158**, 1011–1021 (2014).
29. Kranzusch, P. J. *et al.* Ancient origin of cGAS-STING reveals mechanism of universal 2',3' cGAMP signaling. *Mol. Cell* **59**, 891–903 (2015).
30. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
31. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
32. Terwilliger, T. C. Reciprocal-space solvent flattening. *Acta Crystallogr. D* **55**, 1863–1871 (1999).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).

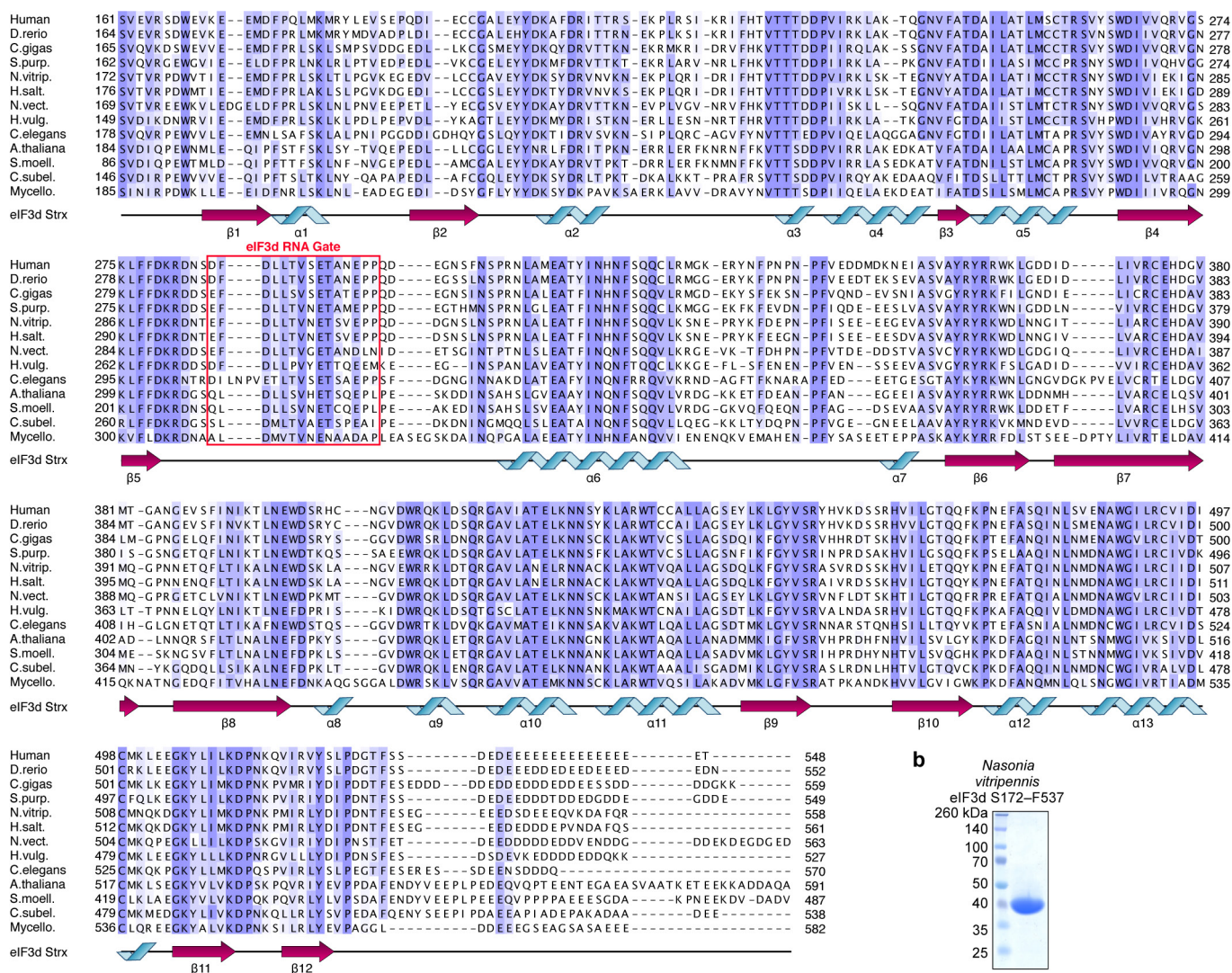


Extended Data Figure 1 | c-Jun expression is unaffected by 4E-BP1 activation. Representative western blot of 293T cells after 24 h treatment with mTOR inhibitor INK128. The results are representative of three independent experiments. For gel source data, see Supplementary Fig. 1.

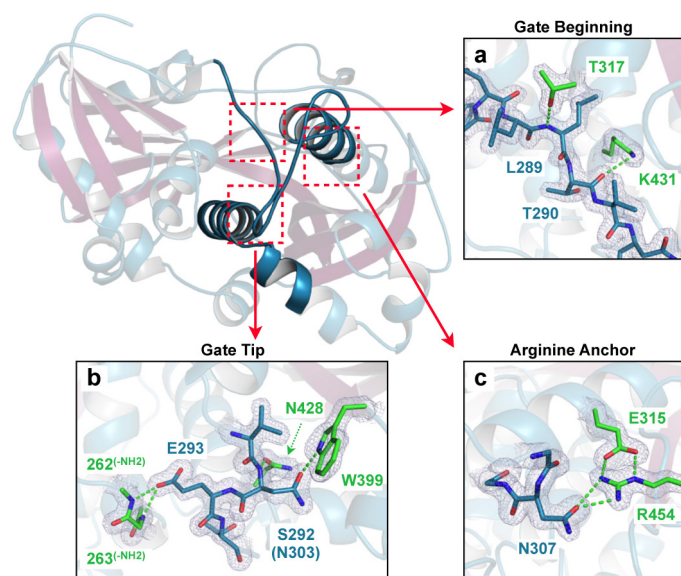


Extended Data Figure 2 | Mapping of a C-terminal region of eIF3d that protects the *c-Jun* 5' cap structure. **a**, Validation of eIF3d subunit identification. eIF3d-cap crosslinking was validated by immunoprecipitation of eIF3d after crosslinking and denaturing the eIF3 complex by boiling in SDS. The result is representative of biological replicates. **b**, Limited proteolysis of eIF3 crosslinked to ³²P-cap-labelled

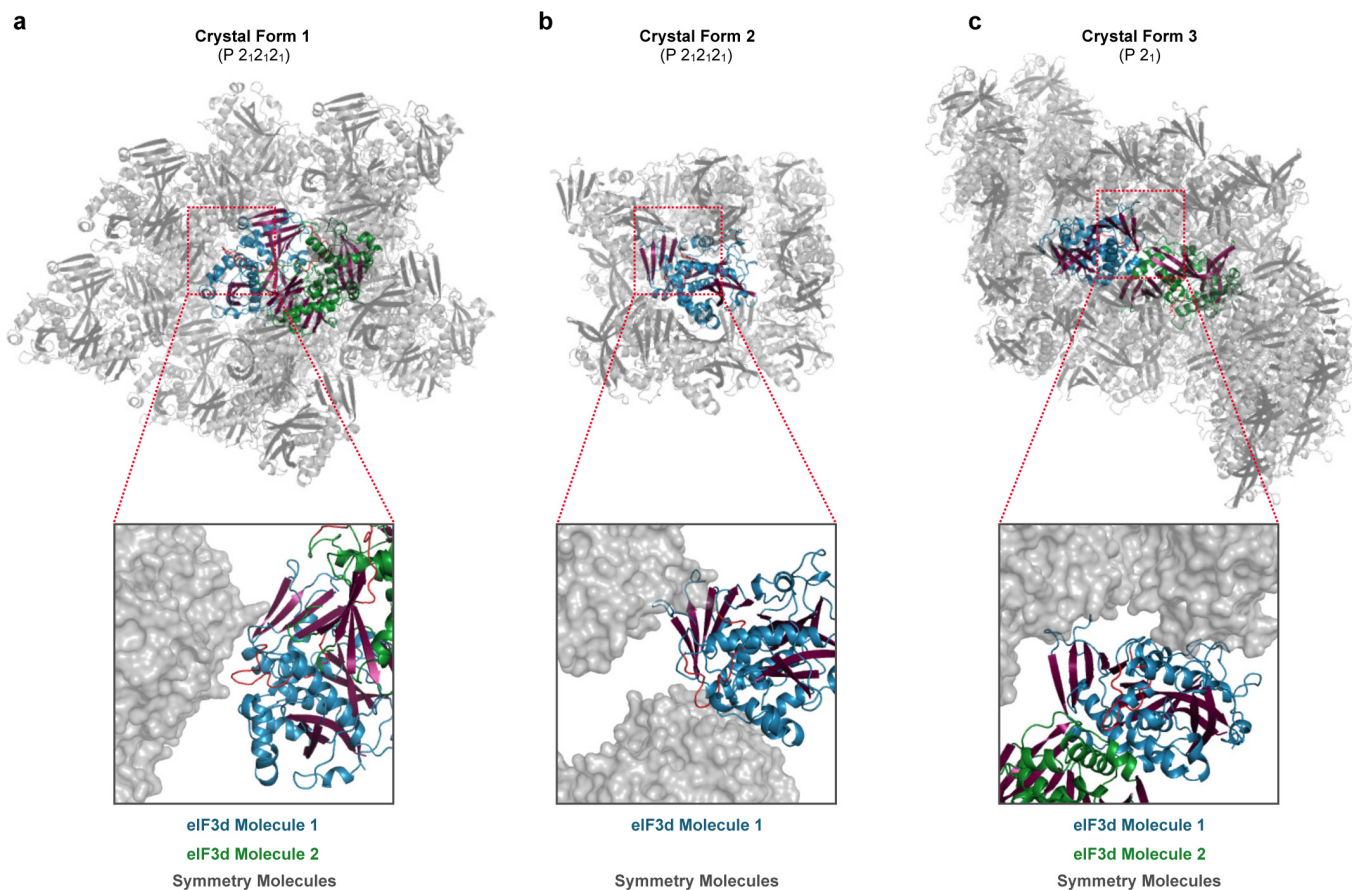
c-Jun 5' UTR RNA. Full-length and proteolysis fragments of eIF3d are indicated by black and maroon arrows, respectively, on the phosphorimage and Coomassie-stained SDS gels. **c**, Mass spectrometry identification of trypsinized peptides from limited proteolysis of cap-crosslinked eIF3d. Identified peptides are highlighted in blue. The results in **b** and **c** are representative of three independent experiments.



Extended Data Figure 3 | Purification of eIF3d cap-binding domain. **a**, Alignment is coloured by phylogenetic conservation of amino acid physicochemical property similarity and a cartoon schematic of the eIF3d secondary structure is depicted below the sequences. Colouring begins at 30% conservation (lightest blue). **b**, Coomassie-blue-stained SDS gel of recombinant *N. vitripennis* eIF3d S172–F537 protein expressed in *E. coli*.

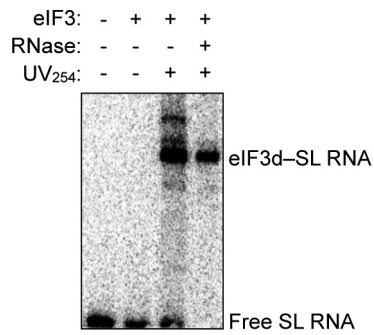


Extended Data Figure 5 | Structural details of eIF3d 'RNA gate' stabilizing interactions. **a–c**, Structural overview of eIF3d with cut-away sections highlighting charged interactions stabilizing the closed 'RNA gate' conformation. No significant van der Waals interactions stabilize the closed gate conformation, supporting likely repositioning of the RNA gate before 5' mRNA cap recognition. Charged interactions occur in three areas: **a**, at the beginning of the gate insertion sequence (gate beginning); **b**, at the tip of the unstructured loop (gate tip); and **c**, at an 'arginine anchor' point stabilizing the return of the loop insertion sequence to the α -helix shared with DXO family endonucleases. Residues are numbered according to the human eIF3d sequence, and all positions are conserved between human and *N. vitripennis* except S292N. eIF3d RNA gate residues are displayed with blue side chains and the residues making stabilizing contacts are coloured in green. $2F_o - F_c$ map regions are shown at 1.5σ .

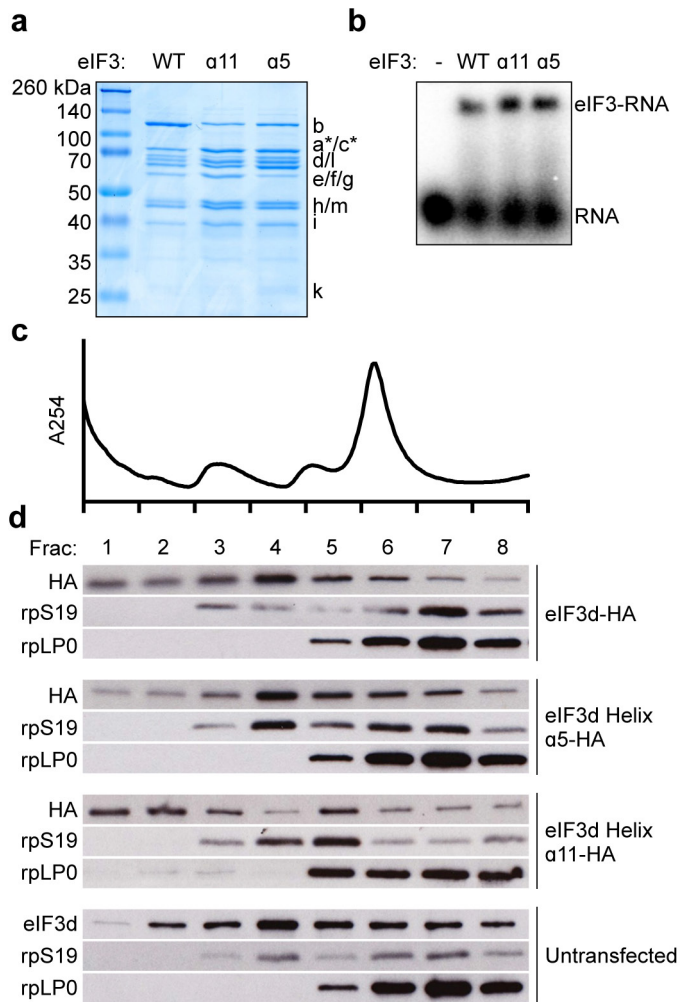


Extended Data Figure 6 | Packing interactions observed in alternative eIF3d crystal forms. Cartoon representation of crystallographic packing in eIF3d crystal form 1, 2 and 3 (**a**, **b** and **c**). Crystal forms 1 and 3 have two copies of eIF3d in the asymmetric unit coloured in blue/magenta and green/magenta, respectively; crystal form 2 has only one copy of eIF3d per asymmetric unit. Symmetry-related molecules are depicted in grey.

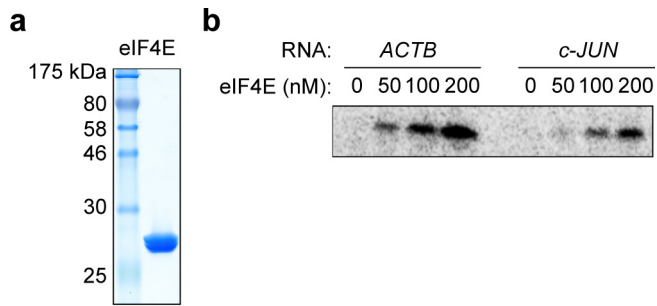
Cut-away zoom illustrates position of the eIF3d RNA gate (red) relative to the nearest symmetry-related molecule. In crystal form 1, the RNA gate is packed against a neighbouring symmetry molecule, but in crystal forms 2 and 3, the RNA gate is positioned towards a major solvent channel. Relative conformation of the RNA gate remains unchanged in either eIF3d crystal form.



Extended Data Figure 7 | eIF3d cap-binding activity requires the eIF3-recruitment stem-loop RNA. Phosphorimage of SDS gel resolving RNase-protected ³²P-cap-labelled *c-Jun* stem-loop RNA crosslinked to eIF3 subunits. The result is representative of three independent experiments.



Extended Data Figure 8 | Incorporation of HA epitope-tagged eIF3d into translation initiation complexes. **a**, Coomassie-blue-stained SDS gel of recombinant eIF3 containing wild-type or helix $\alpha 5$ - or $\alpha 11$ -mutated eIF3d. **b**, Representative native agarose gel electrophoresis of recombinant wild-type and mutant eIF3 complexes bound to the *c-Jun* stem-loop. **c**, Polysome profiles of untransfected 293T cells, plotted as relative absorbance at 254 nm versus elution fractions. **d**, Western blot analysis of eIF3d and the small (rpS19) and large (rpLP0) ribosomal subunits. The results in **b–d** are representative of three independent experiments. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 9 | eIF4E recognizes the 5' end of the *c-Jun* mRNA less efficiently than *ACTB* mRNA. **a**, Coomassie-blue-stained SDS gel of recombinant human eIF4E expressed in *E. coli*. **b**, Phosphorimage of SDS gel resolving RNase-protected ^{32}P -cap-labelled *ACTB* or *c-Jun* 5' UTR RNA crosslinked to eIF4E. The result is representative of three independent experiments. For gel source data, see Supplementary Fig. 1.

Extended Data Table 1 | Summary of data collection, phasing and refinement statistics

	eIF3d Crystal Form 1 (5K4B)	eIF3d Crystal Form 2 (5K4C)	eIF3d Crystal Form 3 (5K4D)	eIF3d (S-SAD)
Data collection				
Space group	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁	P 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.90, 62.93, 192.24	49.01, 61.84, 138.31	49.97, 144.32, 55.30	61.66, 62.72, 193.28
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 109.12, 90.0	90.0, 90.0, 90.0
Wavelength	1.11586	1.11586	1.11587	1.71371
Resolution (Å) ^a	48.06–1.40 (1.42–1.40)	46.20–1.70 (1.73–1.70)	49.13–2.00 (2.05–2.00)	38.66–1.92 (1.96–1.92)
<i>R</i> _{pim}	2.6 (45.3)	5.0 (45.0)	8.6 (52.3)	0.9 (8.1)
<i>I</i> / σ (<i>I</i>)	12.9 (1.5)	9.1 (1.6)	6.5 (1.4)	65.2 (6.7)
<i>CC</i> _{1/2}	99.9 (57.9)	99.6 (63.9)	98.8 (55.6)	100 (94.7)
Completeness (%)	100 (99.2)	99.4 (89.4)	99.8 (98.3)	98.5 (86.1)
Redundancy	15.8 (10.1)	4.1 (3.6)	3.0 (2.7)	129.0 (46.6)
Refinement				
Resolution (Å)	48.06–1.40	46.20–1.70	49.13–2.00	
No. reflections				
Total	2,349,428	193,627	147,613	
Unique	148,455	47,123	49,757	
Free (%)	2	5	5	
<i>R</i> _{work} / <i>R</i> _{free}	17.5 / 19.3	16.2 / 19.6	17.7 / 20.8	
No. atoms				
Protein	5837	2953	5686	
Ligand/ion	3 (Cl)	18 (glycerol)	-	
Water	910	385	741	
<i>B</i> factors				
Protein	21.7	24.2	26.6	
Ligand/ion	31.7	29.1	-	
Water	33.1	36.7	34.3	
r.m.s deviations				
Bond lengths (Å)	0.007	0.012	0.004	
Bond angles (°)	1.167	1.359	0.638	

Single crystals were used to collect data for each structure.

^aValues in parentheses are for highest-resolution shell.

The structural basis of modified nucleosome recognition by 53BP1

Marcus D. Wilson^{1*}, Samir Benlekbir^{2*}, Amélie Fradet-Turcotte^{1†}, Alana Sherker^{1,3}, Jean-Philippe Julien^{2,4,5}, Andrea McEwan¹, Sylvie M. Noordermeer¹, Frank Sicheri^{1,3,4}, John L. Rubinstein^{2,4,6} & Daniel Durocher^{1,3}

DNA double-strand breaks (DSBs) elicit a histone modification cascade that controls DNA repair^{1–3}. This pathway involves the sequential ubiquitination of histones H1 and H2A by the E3 ubiquitin ligases RNF8 and RNF168, respectively^{4–8}. RNF168 ubiquitinates H2A on lysine 13 and lysine 15 (refs 7, 8) (yielding H2AK13ub and H2AK15ub, respectively), an event that triggers the recruitment of 53BP1 (also known as TP53BP1) to chromatin flanking DSBs^{9,10}. 53BP1 binds specifically to H2AK15ub-containing nucleosomes through a peptide segment termed the ubiquitination-dependent recruitment motif (UDR), which requires the simultaneous engagement of histone H4 lysine 20 dimethylation (H4K20me2) by its tandem Tudor domain^{10,11}. How 53BP1 interacts with these two histone marks in the nucleosomal context, how it recognizes ubiquitin, and how it discriminates between H2AK13ub and H2AK15ub is unknown. Here we present the electron cryomicroscopy (cryo-EM) structure of a dimerized human 53BP1 fragment bound to a H4K20me2-containing and H2AK15ub-containing nucleosome core particle (NCP-ubme) at 4.5 Å resolution. The structure reveals that H4K20me2 and H2AK15ub recognition involves intimate contacts with multiple nucleosomal elements including the acidic patch. Ubiquitin recognition by 53BP1 is unusual and involves the sandwiching of the UDR segment between ubiquitin and the NCP surface. The selectivity for H2AK15ub is imparted by two arginine fingers in the H2A amino-terminal tail, which straddle the nucleosomal DNA and serve to position ubiquitin over the NCP-bound UDR segment. The structure of the complex between NCP-ubme and 53BP1 reveals the basis of 53BP1 recruitment to DSB sites and illuminates how combinations of histone marks and nucleosomal elements cooperate to produce highly specific chromatin responses, such as those elicited following chromosome breaks.

To elucidate the molecular basis of 53BP1 recruitment to the chromatin surrounding DNA breaks, we sought to determine the three-dimensional structure of a complex consisting of 53BP1 bound to an NCP modified with H4K20me2 and H2AK15ub. We incorporated a dimethyl-lysine mimic by cysteine alkylation of a histone H4(K20C) mutant¹², yielding H4K_C20me2 (Extended Data Fig. 1a, b). In parallel, we produced H2AK15ub by enzymatic ubiquitination of H2A(K13R and K36R) in the context of a H2A–H2B dimer¹³ (Extended Data Fig. 1c–g). The K36R mutation prevented off-target ubiquitination at this residue (Extended Data Fig. 1c; H2A(K13R and K36R) is hereafter referred to as H2A). Re-purified H4K_C20me2 and H2AK15ub were reconstituted into histone octamers and then wrapped with a 145 base pair fragment of strong nucleosome-positioning ‘601’ DNA¹⁴ to form NCP-ubme (Extended Data Fig. 1h).

The minimal 53BP1 fragment recruited to DSB sites consists of a dimer of the tandem Tudor-UDR segment¹⁰ (residues 1484–1631,

Extended Data Fig. 2a). We used glutathione S-transferase (GST) as a dimerization module to produce a Tudor-UDR dimer (hereafter termed GST–53BP1, Extended Data Fig. 2a). GST–53BP1 bound to NCP-ubme with a K_d of 1 μ M, with only an approximately twofold reduction in association compared to a 53BP1 fragment containing the native oligomerization domain (Extended Data Fig. 2c–e). The resulting GST–53BP1 protein formed a stable complex with NCP-ubme with a 2:1 (53BP1:NCP-ubme ratio) stoichiometry (Fig. 1a and Extended Data Fig. 2e, g).

We determined the structure of the NCP-ubme–GST–53BP1 complex by single particle cryo-EM at 4.5 Å resolution (Fig. 1b, c and Extended Data Fig. 3). The rigid nucleosome forms a symmetrical coin shape, consistent with other NCP–protein structures^{3,15–19}, with extra density attributable to ubiquitin and 53BP1 on each face of the NCP. The densities from the DNA and core secondary structural elements of the histones were unmistakable and similar to previous NCP models^{14,20} (root mean squared deviation (r.m.s.d.) 0.95 Å). The NCP density was sufficiently detailed to model several bulky side-chains (Extended Data Fig. 3g) and the path of the H2A N-terminal tail (see below). Ubiquitin was fit unambiguously into the cryo-EM map, linked via an isopeptide bond to H2A Lys15. No extra density was visualized for the GST moiety, which is probably located near the nucleosome dyad (Extended Data Fig. 3h). The cryo-EM density for the tandem Tudor domain of 53BP1 was weaker than the rest of the structure, but its centre of mass was fixed over the tail of H4, consistent with its tethering to H4K_C20me2 (Fig. 1b).

We attributed extra density projecting from the tandem Tudor domain to residues 1611–1631 of the 53BP1 UDR. This region traverses approximately 50 Å across the NCP, forming contacts with both ubiquitin and histone surfaces (Fig. 1b–d). From its N terminus, the UDR snakes along the NCP-ubme surface, first across a solvent-exposed cleft formed between H4 and H2B, and then through a channel formed between the hydrophobic patch of ubiquitin and the NCP exterior. The UDR continues over the H2B α C helix, over Lys108 and His109, and terminates in a predicted helical conformation above the H2A–H2B acidic patch. Although we were unable to unambiguously assign the UDR density, we deduced the sequence register of the UDR based on complementarity to the local physicochemical environment and by mutational studies described below (Fig. 1d).

We also determined the structure of the unliganded NCP-ubme at 7.7 Å resolution (Extended Data Fig. 4 and Fig. 2a). Comparison of ubiquitin in the bound and unbound states (Fig. 2b, c) suggests that ubiquitin becomes rigidly constrained upon 53BP1 binding following interaction with two separate interfaces: one containing the H2B α C helix and the other involving the UDR of 53BP1 (Fig. 2c, d).

To facilitate validation of the NCP-ubme–GST–53BP1 structure, we substituted enzymatic ubiquitination with chemical ubiquitination

¹The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G1X5, Canada. ²Molecular Structure and Function Program, The Hospital for Sick Children Research Institute, Toronto, Ontario M5G 0A4, Canada. ³Department of Molecular Genetics, University of Toronto, Ontario M5S 3E1, Canada. ⁴Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁵Department of Immunology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁶Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. [†]Present address: Laval University Cancer Research Center, Oncology Axis-CHU de Québec Research Center, Hôtel-Dieu de Québec, Québec City, Québec G1R 2J6 Canada; and Centre Hospitalier Universitaire de Québec Research Center, Université Laval, Québec City, Québec G1R 2J6, Canada.

*These authors contributed equally to this work.

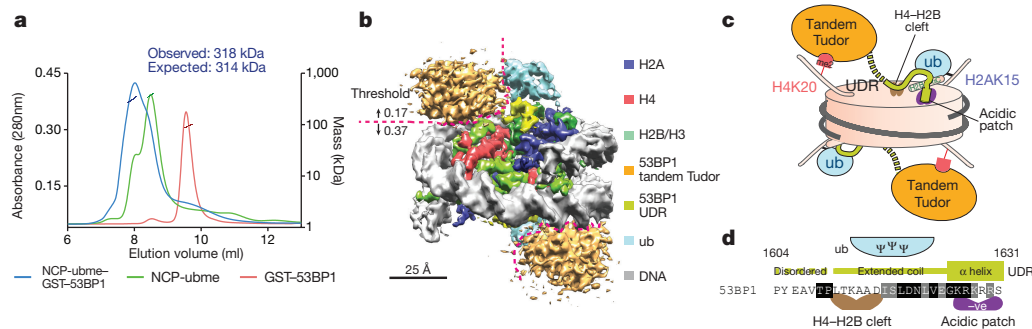


Figure 1 | Architecture of the NCP-ubme-GST-53BP1 complex.

a, Size-exclusion chromatography multi-angle light scattering (SEC-MALS) performed on preformed NCP-ubme-GST-53BP1 complex compared to NCP-ubme or GST-53BP1 alone. **b**, 3D cryo-EM map of the NCP-ubme-GST-53BP1 complex. The map was segmented and coloured according to the respective components. Weaker density

of H2A²¹ using a K15C mutant (Extended Data Fig. 5a–c). We first assessed the relevance of the ubiquitin interaction with the H2B α C helix, which encompasses H2B residues Lys116, Thr119 and Ser123 (Fig. 2d, left and Extended Data Fig. 5d). As predicted, mutation of H2B Thr119 to alanine (H2B(T119A)) or glutamate (H2B(T119E)), a phosphomimetic residue, greatly reduced binding of GST-53BP1 to NCP-ubme (Extended Data Fig. 5e–g). As H2B Thr119 phosphorylation is detected during meiosis²², our results suggest that this histone mark could inhibit 53BP1 recruitment to DSBs. Lys120 of H2B lies in close proximity to the GST-53BP1-bound ubiquitin (Fig. 2d, left). This residue is ubiquitinated during transcriptional elongation and DSB induction^{23–25}, and may antagonize 53BP1 recruitment to chromatin²⁵. We

tested whether 53BP1 binding tolerates H2BK120 ubiquitination and observed that NCPs containing both H2BK120ub and H2AK15ub are efficiently bound by GST-53BP1 (Extended Data Fig. 5h). Therefore it is unlikely that H2BK120 ubiquitination directly blocks 53BP1 recruitment to chromatin.

RNF168 targets both H2A Lys13 and Lys15 in response to DSBs^{7,8}, but only H2AK15ub is recognized by 53BP1 (ref. 10). The NCP-ubme-GST-53BP1 structure hints at the basis for this specificity, which probably involves an interaction between DNA and the N-terminal tail of H2A (Fig. 2d, right). The H2A residues Arg11 and Arg17 straddle the DNA at superhelical location 4.5, placing the Lys15-tethered ubiquitin in a position to interact with the UDR. A similarly ordered H2A tail has been observed previously²⁶, suggesting it can adopt this conformation independently of ubiquitination. Consistent with the H2A-tail conformation underpinning 53BP1 binding specificity, shifting the ubiquitination site by 7 Å to H2AK13 reduced binding to GST-53BP1 (Extended Data Fig. 6a–c), and this binding was restored by a compensatory shift of the two arginine ‘fingers’ in the H2A N-terminal tail sequence (Extended Data Fig. 6c and ref. 10).

The 28-residue UDR, C-terminal to the tandem Tudor domain, confers both ubiquitin and NCP recognition to 53BP1. Cross-linking experiments indicated that the C terminus of the UDR lies near the nucleosome acidic patch, while the N-terminal residues lie near the H2B–H4 cleft (Extended Data Fig. 7a, b). The Sir3 BAH domain also interacts with the H2B–H4 cleft¹⁶, suggesting that this surface may be commonly exploited by NCP-binding proteins (Extended Data Fig. 7c). In addition to H2B-ubiquitin interactions (Fig. 2d, left), direct interactions between the 53BP1 UDR and ubiquitin promote the constrained ubiquitin conformation. However, as the UDR shows no measurable affinity for free ubiquitin¹⁰, these interactions are dependent on UDR binding to the NCP. Ubiquitin presents its hydrophobic patch (comprising Leu8, Ile44 and Val70) towards the middle of the modelled UDR, centred on the aliphatic side chains of Ile1617 and Leu1619 of 53BP1 (Fig. 3a, left), which are also key residues for 53BP1 recruitment to DSB sites¹⁰ (Fig. 3b and Extended Data Fig. 8a, b). The I44A mutation in ubiquitin also reduced binding of GST-53BP1 to NCP-ubme, as observed previously¹⁰ (Extended Data Fig. 8c–e). The Lys6 and His68 side-chains of ubiquitin lie in proximity to two UDR residues, Asp1616 and Asp1620. Mutation of both acidic residues to lysine (D1616K and D1620K) reduced 53BP1 binding to NCP-ubme and impaired recruitment of 53BP1 to DSB sites (Fig. 3b, c), while the K6E and H68E ubiquitin mutations impaired 53BP1 binding to NCP-ubme (Extended Data Fig. 8c–e). Collectively, these results confirm that the unusual mode of ubiquitin recognition by 53BP1 involves the sandwiching of the UDR segment by the NCP surface and the ubiquitin moiety. They also suggest that K6-linked ubiquitin chains, if they occur on H2AK15, are incompatible with 53BP1 binding. However, as most other lysine residues of ubiquitin project

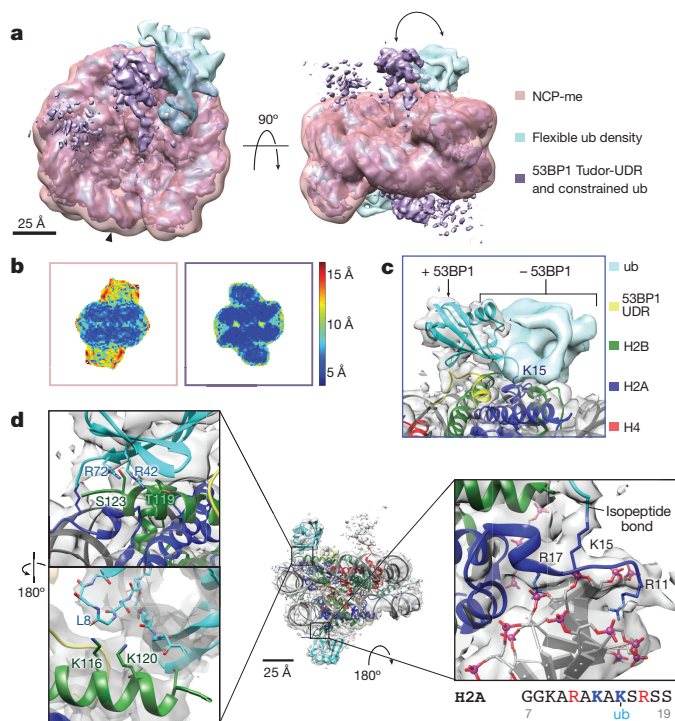


Figure 2 | The molecular basis of H2AK15ub recognition by 53BP1.

a, Overlay of the NCP-ubme and NCP-ubme-GST-53BP1 structures. The cyan density is displayed with a higher threshold. **b**, Equivalent sections through maps of NCP-ubme (left) and NCP-ubme-GST-53BP1 (right), coloured according to local resolution estimates³⁰. **c**, Enlarged view of the region above the H2A Lys15 isopeptide bond with ubiquitin in both structures. **d**, Ubiquitin interactions with the α C helix of H2B (left). Magnified view of the H2AK15ub isopeptide bond, with flanking arginine-finger DNA-interacting residues indicated (right).

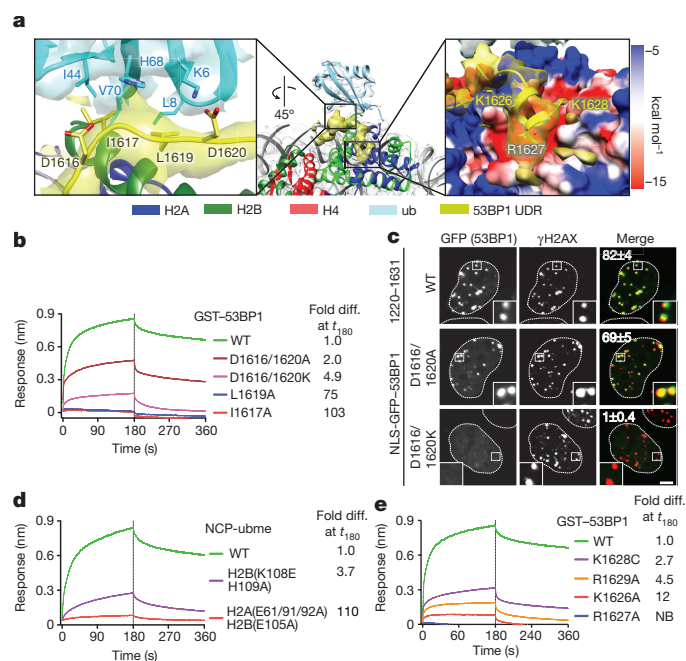


Figure 3 | Multivalent recognition of NCP-ubme by the 53BP1 UDR.

a, UDR-ubiquitin interaction interface (left). Enlarged view of UDR-acidic patch interaction site coloured according to coulombic surface charge (right). **b**, Bio-layer interferometry traces of GST-53BP1 variants at a single concentration binding to immobilized NCP-ubme. WT, wild-type GST-53BP1 protein. **c**, Immunofluorescence analysis of U2OS cells depleted of 53BP1 by siRNA and transfected with GFP-53BP1(1220–1631) variants, 1 h post-irradiation (2 Gy). A representative cell is shown. Scale bar, 5 μ m. Quantification of the percentage of cells with more than 10 GFP (53BP1) foci colocalizing with γ H2AX is shown as mean \pm s.e.m. ($n = 4$). **d**, Bio-layer interferometry traces of GST-53BP1 with NCP-ubme containing proposed UDR-interacting H2A and H2B variants. WT, wild-type histone proteins **e**, Bio-layer interferometry traces of proposed acidic patch-interacting GST-53BP1 UDR variants with NCP-ubme. NB, no detectable binding.

away from the NCP-ubme–GST-53BP1 complex (Extended Data Fig. 8f), the 53BP1–NCP-ubme interaction may be permissive to H2AK15 poly-ubiquitination.

The acidic patch acts as a nexus for protein binding to the nucleosome^{13,15,16,18,19,27,28}. 53BP1 also takes advantage of this surface (Fig. 3a, right and Extended Data Fig. 9a), and its association to NCP-ubme can be outcompeted by the acidic patch-binding LANA peptide¹⁹ (Extended Data Fig. 9b). Removal of the negative charge in the acidic patch, through H2A(E61A, E91A and E92A) and H2B(E105A) mutations, reduced GST-53BP1 binding to NCP-ubme (Fig. 3d and Extended Data Fig. 9c, d). The C-terminal portion of the UDR is highly basic (Fig. 1d) and, in the modelled UDR, residues within the C-terminal α -helix are positioned to interact directly with the acidic patch (Fig. 3a, right). Mutation of any of these residues (K1626A, R1627A, K1628C and K1629A) reduced binding of GST-53BP1 to NCP-ubme (Fig. 3e and Extended Data Fig. 9e, f). In particular, the 53BP1(R1627A) mutation, which also disables recruitment to DSB sites¹⁰, had the most pronounced effect on binding, suggesting that this residue acts as the ‘arginine-anchor’¹⁵ (Extended Data Fig. 9f, g). Mutation of UDR contact elements immediately adjacent to the acidic patch, Lys108 and His109 of H2B (Extended Data Fig. 9a), also decreased the affinity of GST-53BP1 for NCP-ubme (Fig. 3d), emphasizing the role of this interaction surface for 53BP1 binding to chromatin.

The cryo-EM density above the methylated H4 tail was weaker than that of the rest of the NCP-ubme–GST-53BP1 structure (Fig. 4a, b) and did not readily fit the crystal structure of the 53BP1 tandem Tudor domain¹¹. Different structures for this region were obtained from

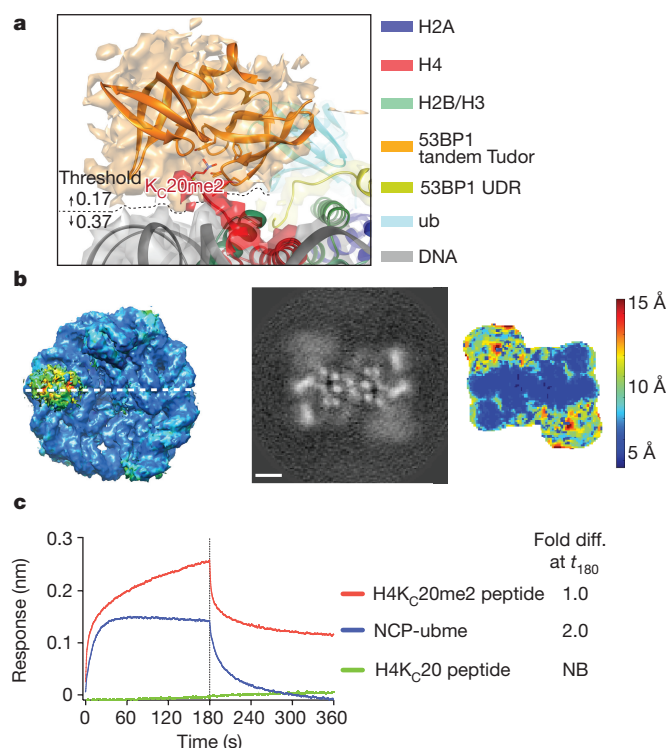


Figure 4 | Flexible association of 53BP1 tandem Tudor domain with NCP-ubme.

a, Magnified view of GST-53BP1 tandem Tudor density over the methylated tail of histone H4. Structure of 53BP1(1484–1601) and H4 peptide (Protein Data Bank accession code 1IG0)¹¹ was fitted. **b**, Indicated sections through the NCP-ubme–GST-53BP1 structure (left), from the calculated map (middle); or coloured according to resolution (right), scale bar, 25 Å. **c**, Bio-layer interferometry traces comparing GST-53BP1(I1617A) binding to NCP-ubme or equal concentrations of H4K_c20 dimethylated or unmethylated peptides. A single GST-53BP1(I1617A) association concentration (2 μ M) is shown.

distinct three-dimensional (3D) classes during image processing, with density in the shape of ‘stem and blossom’ features that we attribute to the site of engagement of H4K_c20me2 (Extended Data Fig. 10a). In all 3D classes, the H4 tail projects upwards, away from the NCP surface, where it engages the methyl-lysine binding pocket of the tandem Tudor domain (Fig. 4a). As the GST-53BP1(I1617A) protein, which is unable to bind to H2AK15ub (Fig. 3b), binds to both a short dimethylated H4 peptide and to NCP-ubme with comparable affinity (Fig. 4c and Extended Data Fig. 10b, c), we surmise that the tandem Tudor domain interacts primarily with a short patch of the H4 tail independently of other nucleosomal elements.

Finally, we tested binding of a 53BP1 fragment that harboured its native oligomerization domain instead of GST to NCP-ubme variants. Each of the elements of NCP-ubme necessary for GST-53BP1 binding is similarly necessary for binding to the native 53BP1 oligomers (Extended Data Fig. 10d). These results indicate that the binding mode described for the GST-53BP1–NCP-ubme structure accurately represents the binding of 53BP1 to nucleosomes.

In summary, the structure of 53BP1 bound to a methylated and ubiquitinated nucleosome highlights how a highly specific biological response, such as the recruitment of 53BP1 to DSB sites, is built from the multivalent engagement of histone post-translational modifications. Far from being the consequence of independent binding events to modified histone residues, the binding of 53BP1 to NCPs requires the coordinated involvement of multiple nucleosomal elements. We also speculate that the intimate association of the 53BP1 dimer with the nucleosome, coupled to its pincer-like binding mode that involves both faces of the nucleosome, may play a key role in stabilizing the nucleosomal barrier that inhibits DNA end resection²⁹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 April; accepted 21 June 2016.

Published online 27 July 2016.

- Lukas, J., Lukas, C. & Bartek, J. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nature Cell Biol.* **13**, 1161–1169 (2011).
- Dantuma, N. P. & van Attikum, H. Spatiotemporal regulation of posttranslational modifications in the DNA damage response. *EMBO J.* **35**, 6–23 (2016).
- Jackson, S. P. & Durocher, D. Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* **49**, 795–807 (2013).
- Thorslund, T. et al. Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage. *Nature* **527**, 389–393 (2015).
- Stewart, G. S. et al. The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell* **136**, 420–434 (2009).
- Doil, C. et al. RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136**, 435–446 (2009).
- Mattioli, F. et al. RNF168 ubiquitinates K13–15 on H2A/H2AX to drive DNA damage signaling. *Cell* **150**, 1182–1195 (2012).
- Gatti, M. et al. A novel ubiquitin mark at the N-terminal tail of histone H2As targeted by RNF168 ubiquitin ligase. *Cell Cycle* **11**, 2538–2544 (2012).
- Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nature Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
- Fradet-Turcotte, A. et al. 53BP1 is a reader of the DNA-damage-induced H2A lys 15 ubiquitin mark. *Nature* **499**, 50–54 (2013).
- Botuyan, M. V. et al. Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
- Simon, M. D. et al. The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007).
- Mattioli, F., Uckelmann, M., Sahtoe, D. D., van Dijk, W. J. & Sixma, T. K. The nucleosome acidic patch plays a critical role in RNF168-dependent ubiquitination of histone H2A. *Nature Commun.* **5**, 3291 (2014).
- Vasudevan, D., Chua, E. Y. & Davey, C. A. Crystal structures of nucleosome core particles containing the ‘601’ strong positioning sequence. *J. Mol. Biol.* **403**, 1–10 (2010).
- McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**, 591–596 (2014).
- Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**, 977–982 (2011).
- Arnaudo, N. et al. The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle. *Nature Struct. Mol. Biol.* **20**, 1119–1121 (2013).
- Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562–566 (2010).
- Barbera, A. J. et al. The nucleosomal surface as a docking station for Kaposi’s sarcoma herpesvirus LANA. *Science* **311**, 856–861 (2006).
- Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
- Long, L., Furgason, M. & Yao, T. Generation of nonhydrolyzable ubiquitin-histone mimics. *Methods* **70**, 134–138 (2014).
- Baarends, W. M. et al. Increased phosphorylation and dimethylation of XY body histones in the *Hr6b*-knockout mouse is associated with derepression of the X chromosome. *J. Cell Sci.* **120**, 1841–1851 (2007).
- Pavri, R. et al. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**, 703–717 (2006).
- Nakamura, K. et al. Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Mol. Cell* **41**, 515–528 (2011).
- Zeng, M. et al. CRL4(Wdr70) regulates H2B monoubiquitination and facilitates Exo1-dependent resection. *Nature Commun.* **7**, 11364 (2016).
- Leung, J. W. et al. Comprehensive structural analysis of mutant nucleosomes containing lysine to glutamine (KQ) substitutions in the H3 and H4 histone-fold domains. *Biochemistry* **50**, 7822–7832 (2011).
- Leung, J. W. et al. Nucleosome acidic patch promotes RNF168- and RING1B/BMI1-dependent H2AX and H2A ubiquitination and DNA damage signaling. *PLoS Genet.* **10**, e1004178 (2014).
- Morgan, M. T. et al. Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**, 725–728 (2016).
- Adkins, N. L., Niu, H., Sung, P. & Peterson, C. L. Nucleosome dynamics regulates DNA processing. *Nature Struct. Mol. Biol.* **20**, 836–842 (2013).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. Szilard for reading the manuscript. We thank M. Forbes for 1D mass spectrometry analysis, C. Davey for providing the ‘601’ DNA, S. Orlicky and D. Ceccarelli for proteins, R. Guenette and E. Strieter for advice on H2A ubiquitination and T. Sixma for discussions. M.D.W. was supported by a Human Frontiers Science Program fellowship. S.M.N. holds a fellowship of the Dutch Cancer Foundation (KWF). J.L.R. is the Canada Research Chair (Tier 1) in Electron Cryomicroscopy and D.D. is a Canada Research Chair (Tier 1) in the Molecular Mechanisms of Genome Integrity. Work was supported by CIHR grants FDN143343 (to D.D.), FDN143277 (to F.S.) and MOP81294 (to J.L.R.), and a Grant-in-Aid from the Krembil Foundation (to D.D.).

Author Contributions M.D.W., F.S. and D.D. initiated the project. M.D.W. assembled protein complexes and performed biochemical assays. S.B. and M.D.W. performed cryo-EM and image analysis. A.F.-T. helped with biochemical assays. A.S. performed cellular assays. J.-P.J. supervised bio-layer interferometry assays. A.M. helped to perform molecular biology. S.M.N. identified *in vitro* ubiquitinated peptides. D.D. supervised the biochemical and cellular experiments and J.L.R. supervised the cryo-EM experiments. M.D.W., F.S., J.L.R. and D.D. wrote the manuscript with input from the other authors.

Author Information Model coordinates for the NCP-ubme–53BP1 UDR structure are deposited in the Protein Data Bank under accession code 5KGF. Cryo-EM maps for NCP-ubme and NCP-ubme–GST-53BP1 are available at the EMDB under codes EMD-8247 and EMD-8246, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.D. (durocher@lunenfeld.ca), J.L.R. (john.rubinstein@utoronto.ca) or F.S. (sicheri@lunenfeld.ca).

Reviewer Information *Nature* thanks Y. Cheng, G. Stewart and G. G. Wang for their contribution to the peer review of this work.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids. Bacterial expression vectors of *Xenopus laevis* histone H3, H3(C110A) and H4(K20C) are described in ref. 10. pET28a synthetic human H2A.1 and pET28a human H2B.1 were a gift from Joe Landry (Addgene plasmids numbers 42634 and 42630, respectively). Expression plasmids for HisTEV-ub pProEx-6 \times His-TEV-ubiquitin and HisTEV-ub G76C pProEx-6 \times His-TEV-ubiquitin G76C were described in refs 10, 31. Plasmids encoding UBA1, UbcH5a and RNF168^{1–113} and pcDNA5–NLS–GFP–53BP1^{1220–1631} were described previously¹⁰. pUC57-8 \times 145 bp Widom-601 was a gift of Curt Davey, Nanyang Technical University.

The human 53BP1 DNA sequence corresponding to residues 1220–1631 was codon-optimized for bacterial expression by GeneArt (ThermoFisher). The native recruitment region, 53BP1 (1220–1631), was amplified and cloned into a modified pETM-30-02 GST-TEV vector described in ref. 10. DNA for the tandem Tudor and UDR segment of 53BP1 (corresponding to residues 1484–1631) was amplified from the codon optimized sequence by PCR and inserted into a modified pET24b, GST-expressing vector with a C-terminal 6 \times His tag, creating the GST–53BP1 expression vector. All mutations described in the manuscript were introduced by site-directed mutagenesis (Quickchange; Stratagene), or via direct DNA synthesis in gBlocks (IDT Technologies) and subsequent sub-cloning. All plasmids were sequenced for verification.

Protein production. RNF168(1–113), UbcH5a, and Uba1 were expressed and purified as described previously¹⁰. In order to remove the 6 \times His tag from RNF168(1–113) and UbcH5a, the pure proteins were cleaved with 6 \times His-TEV protease at room temperature for 4 h in TEV cleavage buffer (50 mM Tris-Cl pH 8, 150 mM NaCl, 2 mM Na-Citrate, 4 mM β -mercaptoethanol) and Ni²⁺ affinity subtracted (HiTrap chelating resin, GE healthcare) to remove 6 \times His-TEV protease and uncleaved 6 \times His tagged protein. Recombinant histones were purified essentially as described³².

GST–53BP1 was produced in *E. coli* BL-21 DE3 CodonPlus cells, after induction with 200 μ M isopropyl β -D-1-thiogalactopyranoside (IPTG) for 18 h at 16 °C. Cells were lysed by sonication and lysozyme treatment in recombinant protein buffer (25 mM Na-phosphate buffer pH 7.4, 300 mM NaCl, 0.1% Triton (v/v), 10% glycerol (v/v), 4 mM β -mercaptoethanol, 1 \times protease inhibitor mix (284 ng ml^{–1} leupeptin, 1.37 μ g ml^{–1} pepstatin A, 170 μ g ml^{–1} phenylmethylsulfonyl fluoride and 330 μ g ml^{–1} benzamidine, 5 μ g ml^{–1} DNaseI). Clarified lysate was applied to a column of Fast-Flow glutathione sepharose (GE healthcare). After extensive washing in high (350 mM NaCl) and low (50 mM NaCl) salt buffers, the protein was eluted with 25 mM reduced glutathione in 25 mM Tris-Cl pH 8, 150 mM NaCl, 10% glycerol (v/v), 4 mM β -mercaptoethanol. The eluate was concentrated with a 30K MWCO centrifugation device (Amicon) and further purified on either S200 10/300 (GE Healthcare) or SEC650 Enrich (BioRad) gel filtration columns in GF buffer (25 mM Tris-Cl pH 7.5, 125 mM NaCl, 2 mM DTT, 5% glycerol (v/v)). All GST–53BP1 variants, including Tudor, UDR mutants and GST-34–53BP1, eluted primarily as dimeric species on size-exclusion chromatography. Protein-containing fractions were concentrated, flash-frozen in liquid nitrogen and stored at –80 °C. For isothermal titration calorimetry (ITC) experiments, proteins were dialysed extensively into ITC buffer (20 mM HEPES pH 7.5, 200 mM KCl, 2 mM Tris (2-carboxyethyl)phosphine (TCEP), 0.5 mM EDTA).

53BP1(1220–1631) was purified by glutathione affinity chromatography as described above. GST-TEV-53BP1(1220–1631) was cleaved from the resin overnight using TEV protease cleavage in TEV cleavage buffer. The eluate was concentrated with a 30K MWCO centrifugation device (Amicon) and further purified on a SEC650 Enrich (BioRad) gel filtration columns in GF buffer. Multiple peaks containing primarily 53BP1(1220–1631) were observed. The main protein containing peak was collected, concentrated, flash-frozen in liquid nitrogen and stored at –80 °C.

HisTEV-ub variants were produced in *E. coli* BL-21 DE3 CodonPlus cells, lysed in recombinant protein buffer, with the NaCl adjusted to 500 mM. Clarified cell lysate was loaded onto a HiTrap chelating column (GE Healthcare) pre-loaded with Ni²⁺ ions. After extensive washing, HisTEV-ub was eluted using a gradient of imidazole and peak-protein containing fractions were concentrated using a 3K MWCO centrifugation device (Amicon). HisTEV-ub was further purified on a S75 10/300 column in GF buffer. Protein-containing fractions were either concentrated by centrifugation or dialysed into water supplemented with 1 mM acetic acid before lyophilization.

All protein concentrations were determined via absorbance at 280 nm using a NanoDrop 8000 (Thermo Scientific), followed by SDS–PAGE and InstantBlue (Expedeon) staining with comparison to known amounts of control proteins.

H4K20 labelling. Cysteine-engineered histone H4(K20C) protein was alkylated essentially as described previously^{12,33}. Briefly, pure histone H4 was

reduced with DTT before addition of a 50-fold molar excess of (2-chloroethyl) dimethylammonium chloride (Sigma-Aldrich). The reaction was allowed to proceed for 4 h at room temperature before quenching with 5 mM β -mercaptoethanol. The H4 protein was separated and desalted using a PD-10 desalting column (GE Healthcare) pre-equilibrated in water supplemented with 2 mM β -mercaptoethanol and lyophilized. Correct incorporation of alkylation agents was assessed by 1D intact weight ESI mass spectrometry (AIMS, Department of Chemistry, University of Toronto; see Extended Data Fig. 1b, additional details below).

Catalytic ubiquitination of H2A. H2A was ubiquitinated within a H2A–H2B dimer, the minimal complex required for K15 specificity¹³. The ubiquitination reaction was performed in ubiquitination reaction buffer (50 mM Tris-Cl pH 7.5, 120 mM NaCl, 5 mM MgCl₂, 1 μ M ZnCl₂, 5 mM ATP, 1 mM DTT) with 120 nM Uba1 (E1), 500 nM UbcH5a (E2), 500 nM RNF168^{1–113} (E3), 12.5 μ M HisTEV-Ub and 4 μ M H2A–H2B at 30 °C for 90 min. The reaction was stopped either by addition of SDS loading buffer or an equal volume of quench buffer (9 M urea, 20 mM Tris-Cl pH 8, 80 mM NaCl, 2 mM EDTA).

The large-scale preparation of catalytically ubiquitinated H2A (Ub–H2A_{CAT}) was purified via loading of the urea-quenched reaction mix on a HiTrap SP column (GE Healthcare) using a peristaltic pump. The loaded column was connected to a FPLC (GE Healthcare) and washed with urea buffer A (7 M urea, 50 mM NaCl, 20 mM Tris-Cl pH 6.8, 2 mM β -mercaptoethanol), and bound histones were eluted using a gradient of urea buffer B (7 M urea, 1 M NaCl, 20 mM Tris-Cl pH 8.8, 2 mM β -mercaptoethanol). HisTEV-H2AK15ub-containing fractions were pooled and enriched over a HiTrap chelating column (GE Healthcare) pre-loaded with Ni²⁺ ions. Extensive washing in urea Ni²⁺ buffer (5.6 M urea, 20 mM Tris-Cl pH 8, 400 mM NaCl, 15 mM imidazole, 2 mM β -mercaptoethanol) removed non-specifically bound unmodified histones. HisTEV-H2AK15ub was eluted in urea Ni²⁺ buffer, supplemented with 300 mM imidazole. The Ni²⁺ column eluate was diluted to a final concentration of 1 M urea, 200 mM NaCl and 55 mM imidazole and the 6 \times His tag at the N terminus of ubiquitin was removed by TEV cleavage overnight at 4 °C. TEV protease and uncleaved HisTEV-H2AK15ub was removed by Ni²⁺ subtraction and the resulting flow through was dialysed extensively in water containing 2 mM β -mercaptoethanol, before lyophilization. H2AK15ub was refolded into histone octamers at an equimolar ratio to the other core histones, essentially as described previously³².

Chemical ubiquitination of H2A. Mutant H2A engineered with one cross-linkable cysteine (H2(K15C)) was ubiquitinated by cross-linking alkylation, essentially as described^{21,34}, with the following modifications. H2A(K15C) (final concentration 700 μ M) was incubated with HisTEV-ub(G76C) (final concentration 350 μ M) in 250 mM Tris-Cl pH 8.6, 8 M urea and reduced with 5 mM TCEP (Sigma-Aldrich) for 30 min at room temperature. The bi-reactive cysteine cross-linker, 1,2-dibromoacetone (DBA, Santa Cruz), was dissolved in dimethyl formamide (DMF) and added to the protein mix to a final concentration of 4.2 mM. The reaction was allowed to proceed on ice for 1 h before quenching with 5 mM β -mercaptoethanol and reducing the pH to pH 8 with trifluoroacetic acid (TFA). Chemically ubiquitinated H2A (H2AK15ub) was purified as described for the catalytic H2A ubiquitination (see also Extended Data Fig. 5a–c). H2AK15ub was refolded into histone octamers at an equimolar ratio to the other core histones, essentially as described previously³².

Nucleosome reconstitution. Nucleosome core particles (NCPs) were reconstituted essentially as described previously^{10,32}. Briefly, the four core histones (with or without modifications) were resuspended in a guanidine hydrochloride denaturing buffer (20 mM Tris-Cl pH 7.5, 7 M guanidine-HCl, 10 mM DTT), mixed at equimolar ratios and then dialysed into a refolding buffer (15 mM Tris-Cl pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM β -mercaptoethanol) to promote folding into a histone octamer. Correctly folded octameric histones complexes were isolated by size exclusion chromatography on a S200 GL 10/300 (GE Healthcare).

Large-scale quantities of pUC57 8 \times 145 bp Widom-601 DNA were isolated using multiple rounds of MaxiPrep kit purifications (Qiagen). Widom 601 145 bp DNA was purified as described previously³² from the pUC57 8 \times 145 bp 601-sequence using EcoRV restriction enzyme to digest the DNA into fragments. Octamers were wrapped using a gradient dialysis technique with Widom-601 145 bp DNA in Rb-low buffer (10 mM Tris-Cl pH 7.5, 200 mM KCl, 1 mM EDTA, 1 mM DTT). Proper assembly of wrapped NCPs, including all H2A and H2B mutants, was analysed by native PAGE as described previously³², and stained with SYBR green dye (ThermoFisher, for example, see Extended Data Fig. 1h). For uncropped original gels, see Supplementary Fig. 1. NCP-ubme was further purified by differential PEG precipitation¹⁷, by incubation with 5% (w/v) PEG-6000 for 10 min on ice, followed by centrifugation at 10,000g at 4 °C for 10 min. The resulting pellet was resuspended in Rb-low buffer.

53BP1–NCP-ubme complex formation. A complex of GST–53BP1 and NCP-ubme was created by incubating the constituent reagents at a 2.5:1 molar ratio.

The complex was purified by differential PEG precipitation, essentially as described¹⁷. Briefly, NCP-ubme (final concentration 20 μ M) was incubated with GST-53BP1 (final concentration 50 μ M) for 10 min on ice. PEG-6000 was added to a final concentration of 5% (w/v) and a precipitate allowed to form for 10 min on ice. The PEG precipitate was centrifuged at 10,000g for 10 min. The NCP-ubme-GST-53BP1 pellet was resuspended in Rb-low buffer and loaded onto a SEC650 Enrich (BioRad) gel filtration column, pre-equilibrated in Rb-low buffer. Fractions eluting earlier than NCP-ubme and GST-53BP1 alone were concentrated to >5 mg ml⁻¹ and stored at 4°C for downstream analysis.

Size-exclusion chromatography multi-angle light scattering (SEC-MALS). SEC-MALS was performed on a Wyatt DAWN TREOS device connected to an Agilent Affinity FPLC using a 30NS prepacked column (Wyatt). The column was pre-equilibrated in Rb-low buffer and 80 μ l of 15 μ M of protein solution (either NCP-ubme, GST-53BP1 or NCP-ubme-GST-53BP1) was autoloading onto the column. Data processing was performed using ASTRA software (Wyatt).

Isothermal titration calorimetry (ITC). ITC measurements were performed on a microCal ITC-200 (GE Healthcare) at 13°C. Prior to the experiment, all proteins were extensively dialysed against ITC buffer. The cell initially contained ~200 μ l of NCP-ubme (H2AK₁₅ub) at a concentration of 23.5 μ M in ITC buffer. GST-53BP1 was added to the syringe at a concentration of 329 μ M (14-fold molar excess of the NCP-ubme cell concentration) and the protein was delivered as two injections of 0.5 μ l, followed by 24 injections of 1.5 μ l, with 3-min intervals between injections. Control experiments were performed under identical conditions to determine enthalpy changes occurring in the cell upon addition of GST-53BP1 to buffer alone. Curve fitting was performed in Origin 7 software (GE Healthcare), using a standard one-site binding model and using the final four data points for baseline subtraction.

Biotin-labelling of NCP-ubme. NCP-ubme complexes were labelled on random surface lysines with biotin by using a 1:2 ratio of NCP to EZ-Link NHS-PEG4-biotin (Thermo-Scientific) in labelling buffer (20 mM HEPES pH 8, 150 mM NaCl, 1 mM DTT, 0.5 mM EDTA). The mixture was incubated for 1 h at room temperature, before buffer exchange in Rb-low to remove unreacted NHS ester using Zeba spin 7K desalting columns (Thermo Scientific). The extent of biotinylation was assessed by binding to streptavidin biosensors (Pall ForteBio), and equal response levels of loading were used in all experiments. Biotinylation did not measurably disassemble NCPs or effect binding to GST-53BP1 (Extended Data Fig. 1h, and data not shown).

Bio-layer interferometry (Octet) assays. Biotinylated NCPs or biotinylated H4 peptides were immobilized on streptavidin biosensors, until reaching a threshold binding of 0.8 nm or 1.5 nm (1.5 nm used for Fig. 4c only), using an Octet RED96 system (Pall ForteBio). Streptavidin biosensors (Pall ForteBio) were pre-equilibrated in Modified Kinetics buffer (1 \times phospho-buffered saline (PBS), 0.025% Tween-20 (v/v), 0.02% (BSA) (w/v), 0.1 mM DTT) for 10 min at room temperature. Experiments were set up in 96-well tray format, using a common protocol: 60 s equilibration in Modified Kinetics buffer, loading of NCP/peptide to threshold value, 60 s wash in Modified Kinetics buffer, 180 s association in GST-53BP1 ligand solution matched to Modified Kinetics buffer, followed by 180 s dissociation in Modified Kinetics buffer. Trays were shaken at 1,000 r.p.m. during the experiment and all experiments were performed at 30°C. Pilot experiments were performed for NCP-ubme and GST-53BP1 mutants to determine the optimal concentrations for kinetic analysis. The assay was performed on four identically loaded sensors, dipped into four wells with consecutive twofold dilutions of ligand (GST-53BP1 variants or 53BP1 (1220-1631)). All dilutions were prepared in Modified Kinetics buffer.

All data were normalized to baseline and subtracted from minimal non-specific binding, based on a blank sensor incubated with the highest concentration of ligand. All data were analysed in Forte Bio Octet analysis software. Full binding kinetics experiments showed a dose-response behaviour correlated with dilution series. Due to the complexity of 53BP1 binding to NCP-ubme, full kinetic fitting was not possible. All experiments were repeated at two immobilization densities, with similar results obtained. Plots of a single ligand concentration (0.5 μ M GST-53BP1 variants in all figures except Fig. 4c) were prepared in Prism (GraphPad). Fold differences of mutants compared to wild-type were determined at a single time point (180 s) across multiple different GST-53BP1 variant concentrations, in the association portion of the binding event.

NCP pull-down assays. Pull-down assays were performed essentially as described¹⁰. Briefly, 4 μ g of GST-tagged 53BP1 constructs were immobilized on glutathione-sepharose resin (GE Healthcare), before incubation with 2.2 μ g of NCP complex in pull-down buffer (50 mM Tris-Cl (pH 7.5), 150 mM NaCl, 1 mM DTT, 0.05% NP-40 (v/v), 0.1% (w/v) BSA). Pull-downs were washed thoroughly in pull-down buffer and resuspended directly in 2 \times SDS loading buffer. For the LANA peptide competition assay, the pull-down was performed as normal, with

the addition of LANA peptide during the incubation with NCP. All pull-down assays were repeated two times, with a single immunoblot displayed.

Immunoblotting. Proteins were separated by SDS-PAGE and transferred to nitrocellulose membranes. All blocking and antibody incubations were performed in Tris-buffered saline containing either 5% (w/v) BSA or 5% (w/v) skimmed milk powder. For western blotting the following commercial primary antibodies were used: rabbit anti-H2B (Abcam), rabbit anti-H3 (Abcam), rabbit anti-GST (Santa Cruz). A rabbit polyclonal H2A antibody was raised against a peptide encompassing human H2A residues 100–130 (KVITIAQGGVLPNIQAVLLPKKTESHKAKGK) coupled to KLH (Covance). Serum from a single immunized mouse serum was found to specifically recognize histone H2A (validation of the antibody is the Supplementary Fig. 2). HRP-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch) secondary antibodies were used with enhanced chemiluminescence solution (ECL supersignal, Thermo Scientific) was used for protein detection. All pull-down assays were repeated two times, with a single immunoblot displayed. For uncropped original gels, see Supplementary Fig. 1.

Protein cross-linking. NCP-ubme with H2B cysteine mutations (N84C or E105C) were assembled as previously described, except using a H3 where the sole native cysteine was mutated (H3(C110A)). NCPs were desalted using Zeba spin desalting columns (Thermo Scientific) to remove reducing agents into degassed cross-linking buffer (20 mM Tris-Cl (pH 6.8), 100 mM NaCl, 1 mM EDTA). NCPs were then incubated in a 1:5 ratio with the bifunctional maleimide bismaleimidoethane (BMOE) for 1 h at room temperature. BMOE-conjugated NCPs were again desalted and quantified. A final concentration of 1.5 μ M BMOE-NCP-ubme was mixed with a sixfold molar excess of freshly desalted GST-53BP1 cysteine mutants (T1609C or K1628C). The reaction was incubated for 2 h at room temperature and quenched by addition of SDS loading buffer supplemented with 25 mM β -mercaptoethanol. The extent of crosslinking was assessed by immunoblotting with anti-H2B antibodies.

Mass spectrometry. *In vitro* ubiquitination reaction products of H2A and H2B (4 μ g of each) were separated on a 15% SDS-PAGE gel and bands were excised at the height of the non-modified, mono- and di-ub forms (based on molecular size markers run in parallel). Proteins were in-gel digested at 37°C for 330 min using 100 ng trypsin (Worthington), and gel-extracted using TFA-acidification and acetonitrile dehydration. Peptides were cleaned up using C18 stage tips (Thermo Scientific) and dried to completeness. Peptides were reconstituted in 5% (v/v) formic acid and loaded onto a 12 cm fused silica column with a pulled tip that was packed in-house with 3.5 μ m Zorbax C18 (Agilent Technologies). Samples were analysed using an Orbitrap Elite (Thermo Scientific) coupled to an Eksigent nanoLC ultra (AB SCIEX). Peptides were eluted from the column using a 90 min linear gradient from 2% to 35% (v/v) acetonitrile in 0.1% (v/v) formic acid. Tandem MS spectra were acquired in a data-dependent mode for the top 10 most abundant multiply charged peptides, with a dynamic exclusion duration of 20 s. Tandem MS spectra were acquired using collision-induced dissociation. Mascot was used to search spectra against the human Refseq_V53 database, allowing up to two missed cleavages and including GlyGly (K) and LeuArgGlyGly (K) as variable modifications. Fragmentation spectra of diGly-peptides identified by Mascot were manually verified.

Determination of the intact mass of histone H4 peptides and H4 protein was performed at AIMS Mass Spectrometry facility, Department of Chemistry, University of Toronto. Electrospray ionization (ESI) mass spectra were acquired in positive ion mode using a 6538 UHD model quadrupole time-of-flight mass spectrometer equipped with a 1260 Infinity Series HPLC (Agilent Technologies, Santa Clara, CA). Intact proteins were mass analysed following online de-salting using a Tricorn 5/50 column packed with Sephadex G-25 size-exclusion media (GE Healthcare). The mobile phase was 1:1 (v/v) 0.1% aqueous formic acid: methanol, flowing at a rate of 500 μ l min⁻¹. Samples were diluted in mobile phase to a concentration of approximately 0.1 μ M and injections on the column were 2.5 μ l.

Peptides. The wild-type LANA₁₋₂₃ peptide (Biotin-LC-MAPPGMR LRSRSTGAPLTRGSY), the non-binding LANA₁₋₂₃ 8LRS10 (Biotin-LC-MAPPGMR AAGRSTGAPLTRGSY) and the biotinylated H4K20C₁₂₋₂₇ peptide (Biotin-LC-KGGAKRHRCVLRDNIQ) were synthesized by BioBasic. Peptides were modified to create dimethyl-lysine analogues (using (2-chloroethyl)-dimethylammonium chloride (Sigma-Aldrich)) or lysine analogues (using (2-bromoethyl)-ammonium bromide (Millipore)) as described for histone H4. Peptides were purified from the reactant materials by SepPak C18 columns (Cell Signaling) and lyophilized dry, before resuspension in DMSO for downstream assays. Correct incorporation of alkylation agents was assessed by intact weight ESI Mass spectrometry (AIMS, Chemistry Department, University of Toronto, Extended Data Fig. 10b, c).

Cryo-EM grid preparation and microscopy. Holey carbon film-coated EM grids were prepared with arrays of 500–800 nm holes by nanofabrication³⁵. 2.5 μ l of NCP-ubme or NCP-ubme-GST-53BP1 complex was diluted to a final salt

concentration of ~ 50 mM. The low-salt complexes were applied to grids and allowed to equilibrate for 5 s in a FEI Vitrobot grid preparation robot, and blotted from both sides for 15 s before freezing in a liquid ethane/propane mixture (1:1 v/v)³⁶. Grids were subsequently stored in liquid nitrogen, before transfer to a Gatan 626 cryotransfer specimen holder. Samples were imaged with a FEI F20 electron microscope, equipped with a field emission gun and operating at 200 kV. Movies were acquired with a Gatan K2 Summit direct detector device (DDD) camera using a calibrated magnification of $34,483\times$, resulting in a physical pixel corresponding to 1.45 \AA . The DDD was used in counting movie mode with 5 electrons per pixel per s for 15 s and 0.5 s per frame. This exposure rate resulted in 1.2 electrons per \AA^2 per frame and a total exposure of 36 electrons per \AA^2 on the specimen. A total of 227 movies were acquired for the NCP-ubme with a defocus range from 1 to $3.2\text{ }\mu\text{m}$. 319 movies were acquired for the NCP-ubme–GST–53BP1 complex with defocus between 0.8 and $4.1\text{ }\mu\text{m}$. NCP-ubme, and NCP-ubme–GST–53BP1 data sets were treated similarly.

Cryo-EM image analysis. Individual frames in a movie stack were aligned and averaged using the programs `alignframes_lmbfsgs` and `shiftframes`³⁷. Contrast transfer function (CTF) parameters were calculated from the averaged frames using `CTFFIND3`³⁸. Manual inspection of micrographs and their corresponding power spectra was performed in Relion 1.3 (ref. 39) and undesirable micrographs were discarded due to contamination or lack of high resolution Thon rings in their power spectrum. Automatic particle picking, based on manually selected templates, was performed in Relion 1.3. After particle extraction, beam induced particle motion between frames was corrected with `alignparts_lmbfsgs`³⁷. A previously measured 2% magnification anisotropy was corrected as described⁴⁰. Extracted particles were subject to 2D classification in Relion 1.3 and classes with averages that resembled the expected projections of NCPs were selected for 3D classification. A low-pass filtered model of NCP based on Protein Data Bank accession code 1KX5 (ref. 20) was used as a template for 3D classification into 9 classes (NCP-ubme and GST–53BP1) or 4 classes (NCP-ubme). Particle images from 3D classes that showed high-resolution features were refined further. Refined maps of NCP-ubme–GST–53BP1 were sharpened in Relion 1.3 with an automatically determined B-factor. NCP-ubme maps were not sharpened. Global resolution estimates were determined using the $\text{FSC} = 0.143$ criterion after a gold-standard refinement⁴¹. Local resolution was estimated with `ResMap`³⁰. Calculations with Relion 1.3 were performed using the SickKids High Performance Facility (Hospital for Sick Children, Toronto). All programs used are freely available through the respective cited distributors.

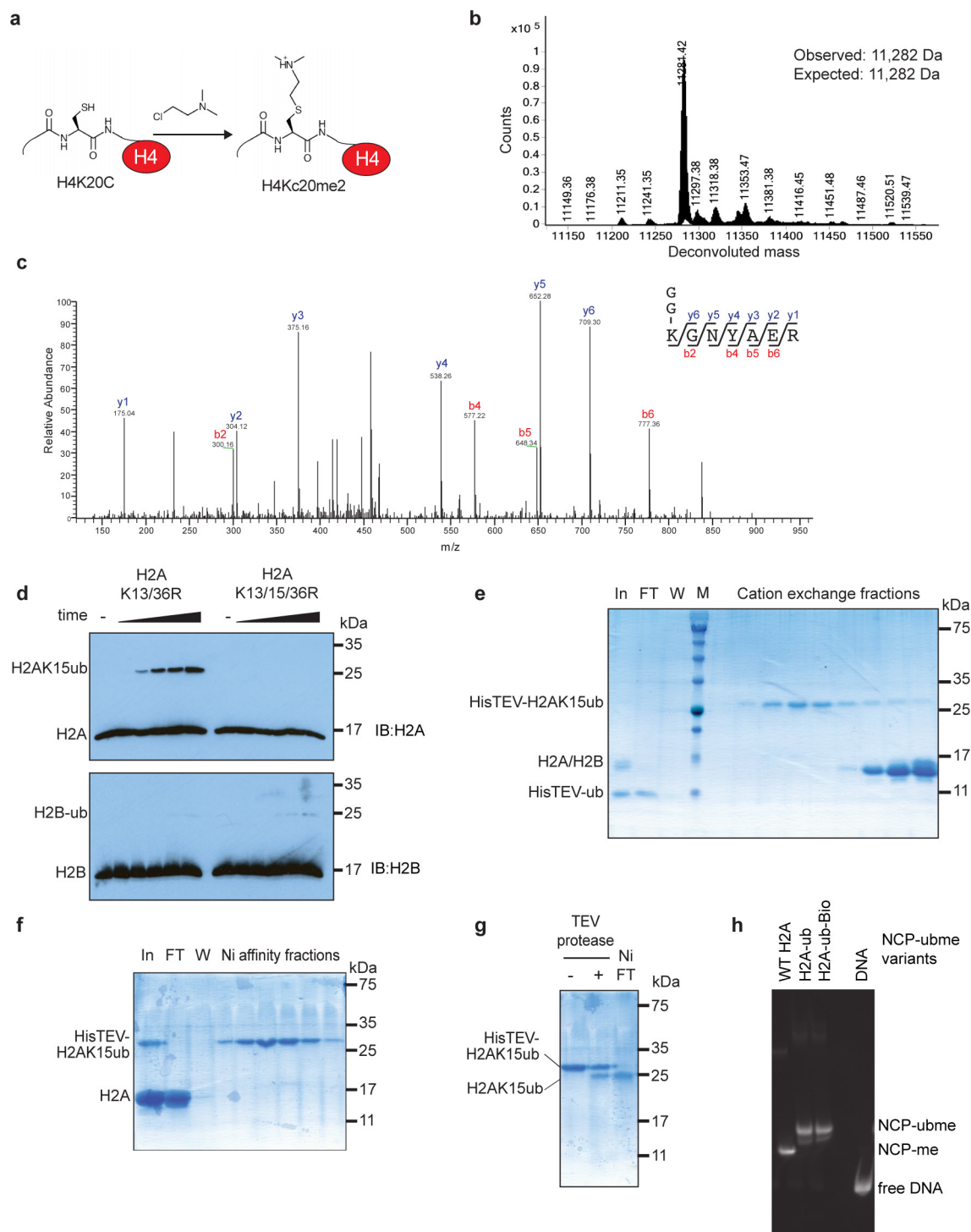
Structure editing and modelling. The atomic models of Widom-601 DNA (Protein Data Bank accession code 3LZ0)¹⁴, octameric histones (Protein Data Bank accession code 1KX5)²⁰, ubiquitin (Protein Data Bank accession code 1UBI)⁴² and H4K20me2/53BP1 tandem Tudor domain (Protein Data Bank accession code 2IG0)¹¹ were fitted in to the 3D maps using UCSF Chimera⁴³ without allowing flexibility. Map segmentation was performed in UCSF Chimera. In Fig. 2a, density corresponding to ubiquitin in the NCP-ubme structure was displayed with a threshold of 0.125. The rest of the NCP-ubme and NCP-ubme–GST–53BP1 are displayed with a threshold of 0.29 and 0.395, respectively. The H2A/H2B sequence was mutated to the human H2A(K13R and K36R) and H2B manually in UCSF Chimera. A polyaniline model of the UDR was built within the UDR density in Coot⁴⁴, which compared well to predicted structures generated by Robetta⁴⁵. The UDR model was mutated and fitted in UCSF Chimera, followed by iterative rounds of real-space refinement in PHENIX⁴⁶ and model optimization in Coot. All figures were prepared in UCSF Chimera.

Cell culture. U-2-OS (U2OS) cells were purchased from ATCC and verified mycoplasma free. Cells were cultured with McCoy's Medium (Gibco) supplemented with 10% fetal bovine serum and maintained at 37°C in 5% CO_2 atmosphere

conditions. Knock-down by single duplex siRNA to 53BP1 (ThermoFisher, D-003548-01, target sequence: 5'-GAGAGCAGAUGAUCCUUUA-3') was performed with RNAiMAX (Invitrogen) two days before fixation, following the manufacturers' instructions. One day before fixation, cells were transiently transfected with the plasmid NLS–GFP–53BP1(1220–1631) DNA using Lipofectamine-3000 transfection reagent (Invitrogen).

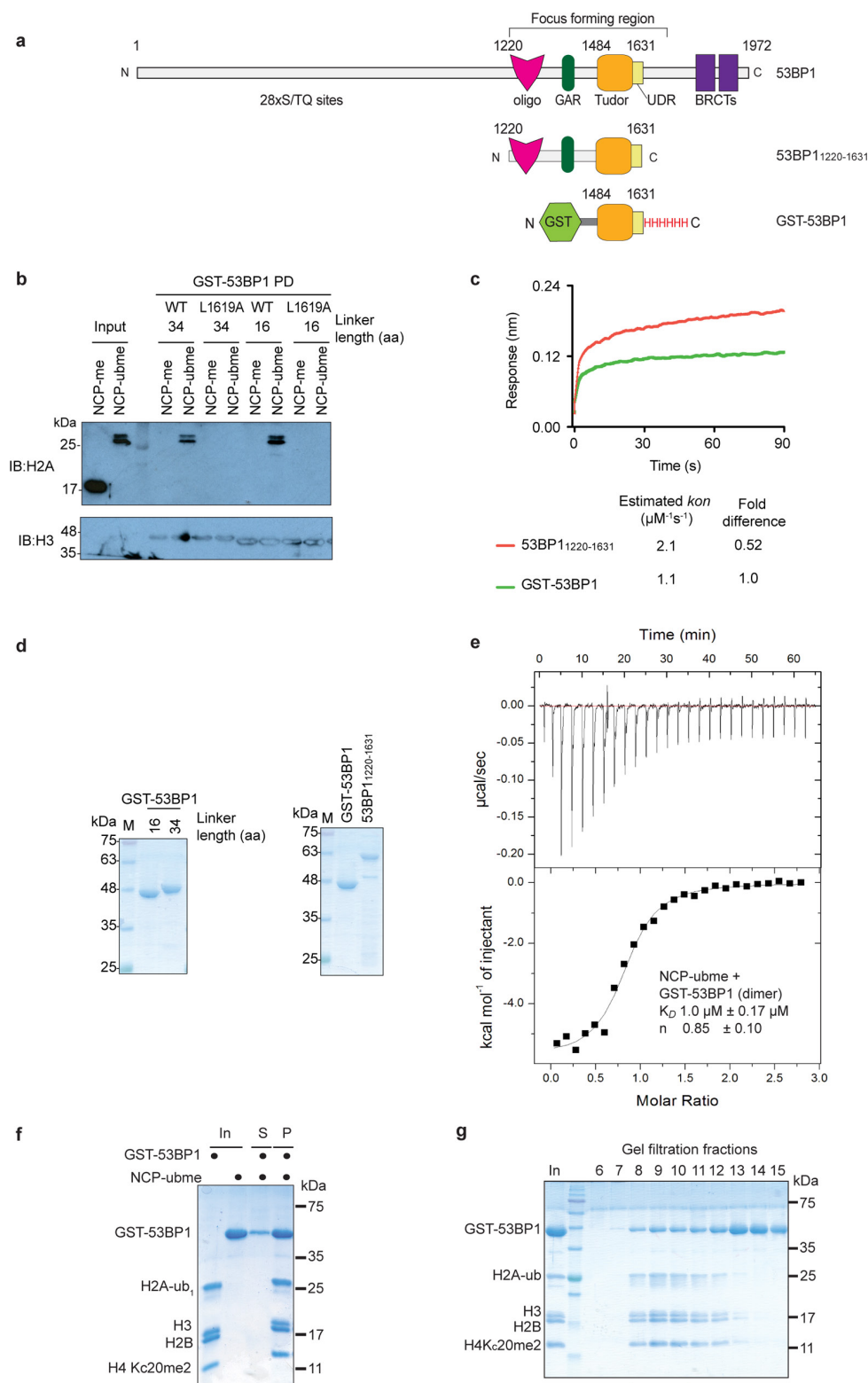
Immunofluorescence microscopy. One hour after exposure to 2 Gy of ionizing radiation cells were fixed with 4% (w/v) formaldehyde. Fixed cells were permeabilized in PBS, 0.3% (v/v) Triton X-100 and blocked in blocking buffer (PBS, 10% (v/v) goat serum, 0.5% (v/v) NP-40, 0.5% (w/v) Saponin). Coverslips were stained with anti- γH2AX (Millipore) and anti-mouse IgG Alexa–fluor 647 (Millipore) secondary antibody. DNA was counterstained with DAPI, which was used to trace the outline of nuclei. Stained cells were visualized on a LSM780 Zeiss confocal microscope. Quantification was performed on 100 U2OS cells ($n = 3$), in which a cell with >10 GFP–53BP1 foci was considered positive.

31. Juang, Y. C. *et al.* OTUB1 co-opts Lys48-linked ubiquitin recognition to suppress E2 enzyme function. *Mol. Cell* **45**, 384–397 (2012).
32. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
33. Simon, T. W. *et al.* The use of mode of action information in risk assessment: quantitative key events/dose-response framework for modeling the dose-response for key events. *Crit. Rev. Toxicol.* **44** (Suppl 3), 17–43 (2014).
34. Orthwein, A. *et al.* A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **528**, 422–426 (2015).
35. Marr, C. R., Benlekhir, S. & Rubinstein, J. L. Fabrication of carbon films with $\sim 500\text{nm}$ holes for cryo-EM with a direct detector device. *J. Struct. Biol.* **185**, 42–47 (2014).
36. Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
37. Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *J. Struct. Biol.* **192**, 188–195 (2015).
38. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
39. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Zhao, J., Brubaker, M. A., Benlekhir, S. & Rubinstein, J. L. Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. *J. Struct. Biol.* **192**, 209–215 (2015).
41. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
42. Ramage, R. *et al.* Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J.* **299**, 151–158 (1994).
43. Pettersen, E. F. *et al.* UCSF Chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
45. Kim, D. E., Chivian, D. & Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
47. Rufer, A. C., Thiebach, L., Baer, K., Klein, H. W. & Hennig, M. X-ray structure of glutathione S-transferase from *Schistosoma japonicum* in a new crystal form reveals flexibility of the substrate-binding site. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **61**, 263–265 (2005).
48. Weeks, S. D., Grasty, K. C., Hernandez-Cuevas, L. & Loll, P. J. Crystal structures of Lys-63-linked tri- and di-ubiquitin reveal a highly extended chain architecture. *Proteins* **77**, 753–759 (2009).



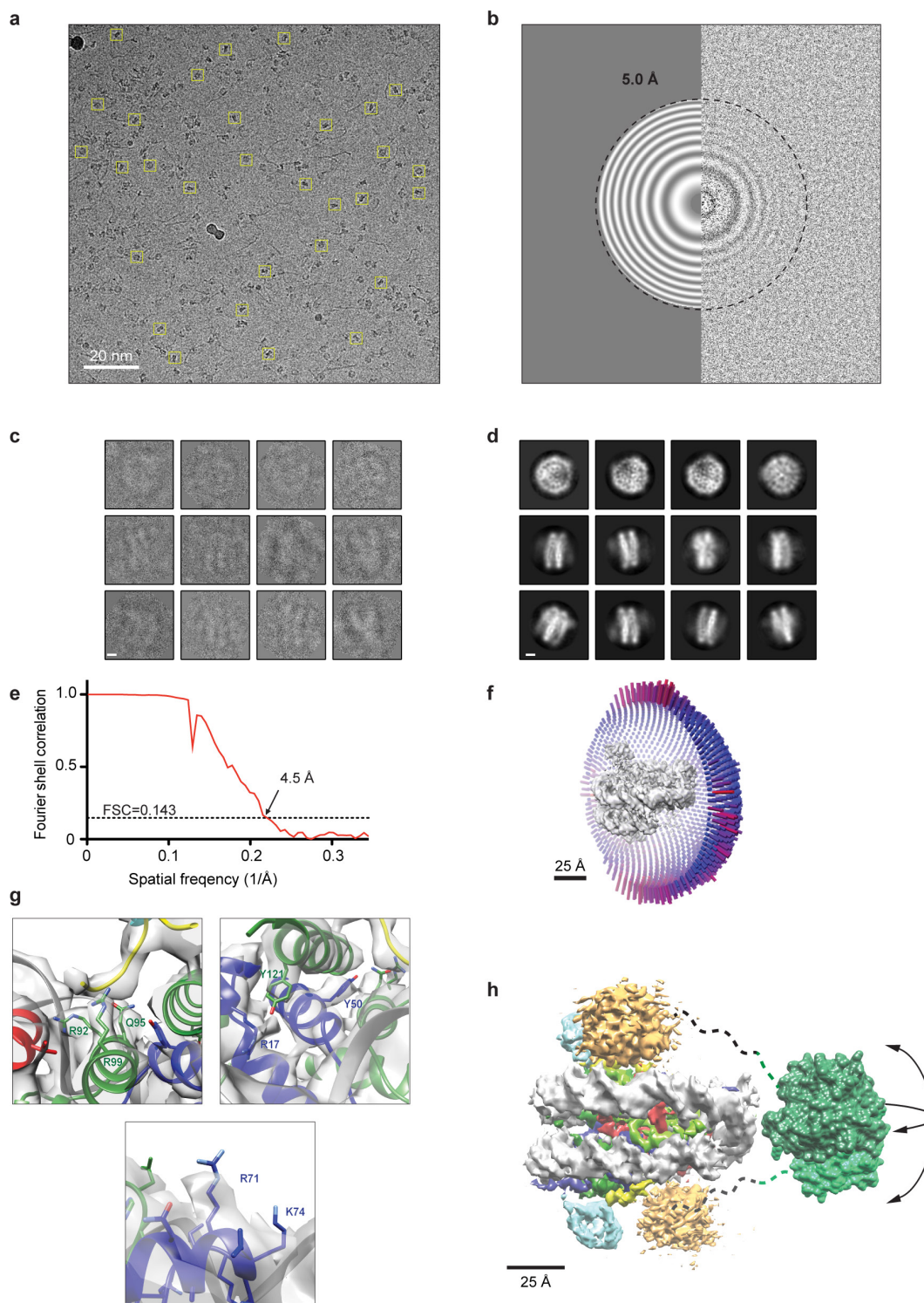
Extended Data Figure 1 | Generation of homogeneously methylated and ubiquitinated NCPs. **a**, Schematic of H4 cysteine alkylation to create a dimethyl-lysine analogue. **b**, 1D intact mass spectra of the alkylated H4Kc20me2 protein after desalting and lyophilization. **c**, Mass spectrum of the identified off-target K36 ubiquitination. Fragmentation spectrum of the 37-K(GlyGly)GNYAER-43 H2A peptide (476.237091 Da, 2^+ charge state, Mascot ions score: 45). This spectrum originates from di-ubiquitinated forms of *in vitro* ubiquitinated H2A that were separated by SDS-PAGE, subjected to limited trypsin digestion and analysed by tandem mass spectrometry. **d**, Immunoblot analysis of a H2A-H2B dimer ubiquitination reaction. Comparison of K13R, K15R and K36R triple mutated and K13R and K36R double mutated H2A variants using optimized conditions to minimize off-target ubiquitination for large-scale reactions. **e**, SDS-PAGE analysis of the first step in H2AK15ub purification, cation exchange chromatography (in, input;

FT, flow-through; W, wash; M, molecular weight marker). The ubiquitin is tagged with an N-terminal hexa-histidine tag and TEV cleavage site (termed HisTEV-ub). **f**, SDS-PAGE analysis of the second step in H2AK15ub purification, nickel ion affinity chromatography. **g**, SDS-PAGE analysis of TEV protease cleavage of HisTEV-H2AK15ub and subsequent nickel column depletion. Cleaved H2AK15ub flows through (Ni FT) the column, while uncleaved product and His-tagged TEV protease binds it (not shown). **h**, Native polyacrylamide gel analysis of wrapped NCPs. The gel was stained with SYBR green to identify DNA. Wrapping of NCPs results in quenching of the SYBR green signal and a shift in the electrophoretic mobility of the DNA. Ubiquitinated NCPs (H2A-ub) appear as a doublet, which runs higher than solely methylated NCPs (WT H2A). Biotinylation on NCP surface lysines, required for downstream bio-layer interferometry analysis, does not measurable effect migration in the gel (H2A-ub-Bio). WT, wild-type H2A protein.



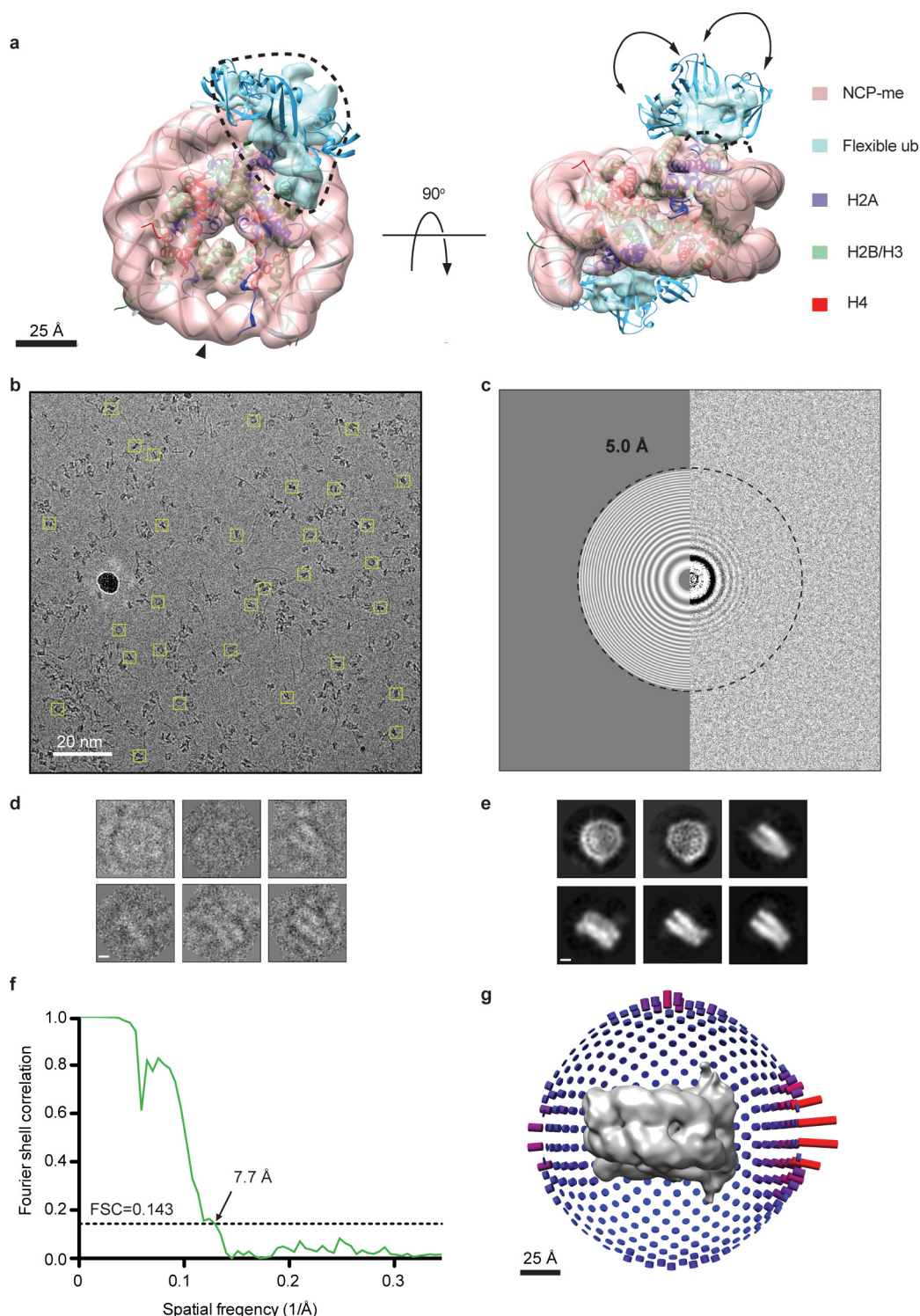
Extended Data Figure 2 | Formation of NCP-ubme and NCP-ubme-GST-53BP1 complexes. **a**, Schematic representation of full-length human 53BP1, 53BP1 with the native recruitment region, 53BP1(1220–1631), and GST-53BP1(1484–1631) (termed GST-53BP1). 53BP1(1220–1631) and GST-53BP1 constructs are used throughout this manuscript. Identified domains are highlighted; oligo: oligomerisation domain. **b**, Pull-down assay of GST-53BP1 variants containing either the 16-residue linker used throughout this study or a longer 34-residue linker. The L1619A UDR mutant was included as a negative control. **c**, Bio-layer interferometry assays of GST-53BP1 or a native 53BP1(1220–1631) fragment. **d**, SDS-PAGE of purified 53BP1 proteins used in the pull-down and

bio-layer interferometry assays. **e**, Isothermal titration calorimetry (ITC) measurement investigating the affinity of GST-53BP1 (syringe) to NCP-ubme (cell). Data reported as the mean \pm s.e.m. ($n = 2$). **f**, SDS-PAGE analysis of NCP-ubme-GST-53BP1 complex formation by differential PEG precipitation¹⁷ (in, input; S, soluble supernatant; P, pellet). An excess of GST-53BP1 was added, which was not precipitated with the NCP-ubme. **g**, SDS-PAGE analysis of size exclusion chromatography fractions, isolating NCP-ubme-GST-53BP1. In, input; M, size markers. Fractions 8–10 were pooled and used for SEC-MALS analysis (Fig. 1a) and subsequent structure determination.



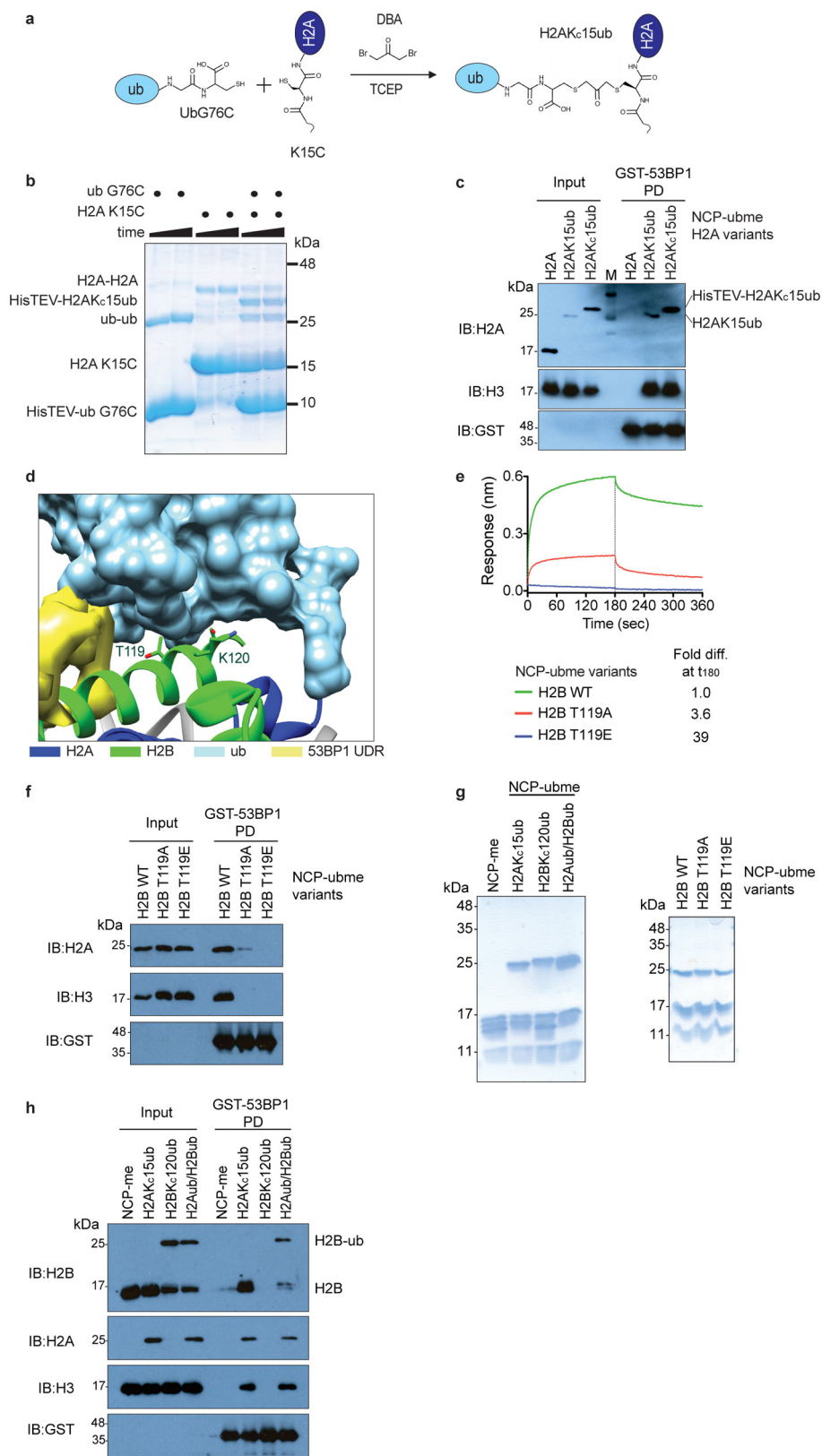
Extended Data Figure 3 | Cryo-EM structure determination and validation of NCP-ubme-GST-53BP1 complex. **a**, Representative cryo-EM micrograph of the NCP-ubme-GST-53BP1 complex. Example particle images in different orientations are boxed. **b**, Power spectrum from a representative micrograph showing Thon rings that extend beyond 7 Å resolution. **c**, Example particle images, scale bar corresponds to 25 Å. **d**, Examples of 2D class averages obtained during image processing of the NCPubme-GST-53BP1 complex (CTF corrected, inverted contrast). Scale bar, 25 Å. **e**, Fourier shell correlation curve after a gold-standard map refinement. **f**, Euler angle distribution plot of all particles used for the symmetrized final map. Bar length and colour (blue, low; red, high) corresponds to number of particle images contributing to each view. **g**, Magnified view of the H2B-H4 cleft with clear side chain density observed for H2B Arg89, Gln92 and Arg96 (top left). Magnified view of

the H2B α C helix and H2A α 1 helix (top right). Densities for aromatic side-chains of H2A Tyr50, H2B Tyr121 and the bulky residue H2A Arg17 are visible. Magnified view of the C terminus of the H2A α 2 helix, with density for base of the side chain of Arg71 visible (bottom). **h**, Schematic of predicted location of flexible GST moiety (green) used for 53BP1 dimerization (Protein Data Bank accession code 1Y6E)⁴⁷. No cryo-EM density can be attributed to GST, suggesting that it is highly flexible between different particles in the population sampled. Dashed lines indicate the 16 amino acid linker region incorporated in the GST-53BP1 construct (black) and the flexible C-terminal tail of the GST (green). The linker peptide region could span up to ~80 Å, allowing substantial flexibility of the GST dimers, shown here positioned ~50 Å from the N terminus of the modelled Tudor domain.



Extended Data Figure 4 | Cryo-EM structure determination and validation of NCP-ubme complex. **a**, Surface rendering of the NCP-ubme complex, viewed along the DNA axis and the orthogonal direction. Density corresponding to ubiquitin was segmented, Gaussian filtered and displayed with a threshold of 0.125 (area within dashed line). The rest of the NCP-ubme is displayed with a threshold 0.35. Rigid body fitting of the high resolution structure of histone octamers (Protein Data Bank accession code 1KX5)²⁰, Widom-601 145bp DNA (Protein Data Bank accession code 3LZ0)¹⁴ and ubiquitin (Protein Data Bank accession code 1UBI)⁴² into NCP-ubme density is shown. Ubiquitin could not be readily placed in the attributed density. **b**, Representative cryo-EM

micrograph of the NCP-ubme complex. Example particle images showing different orientations are boxed. **c**, Power spectrum from a representative micrograph showing Thon rings. **d**, A selection of particles images after extraction from the data set. Scale bar, 25 Å. **e**, Examples of 2D class averages obtained during image processing (CTF corrected, inverted contrast); scale bar, 25 Å. **f**, Fourier shell correlation curve after gold-standard map refinement. **g**, Euler angle distribution plot of all particle images used for the symmetrized final map. Bar length and colour corresponds to number of particle images in each view that contributed to final 3D map (blue, low; red, high).

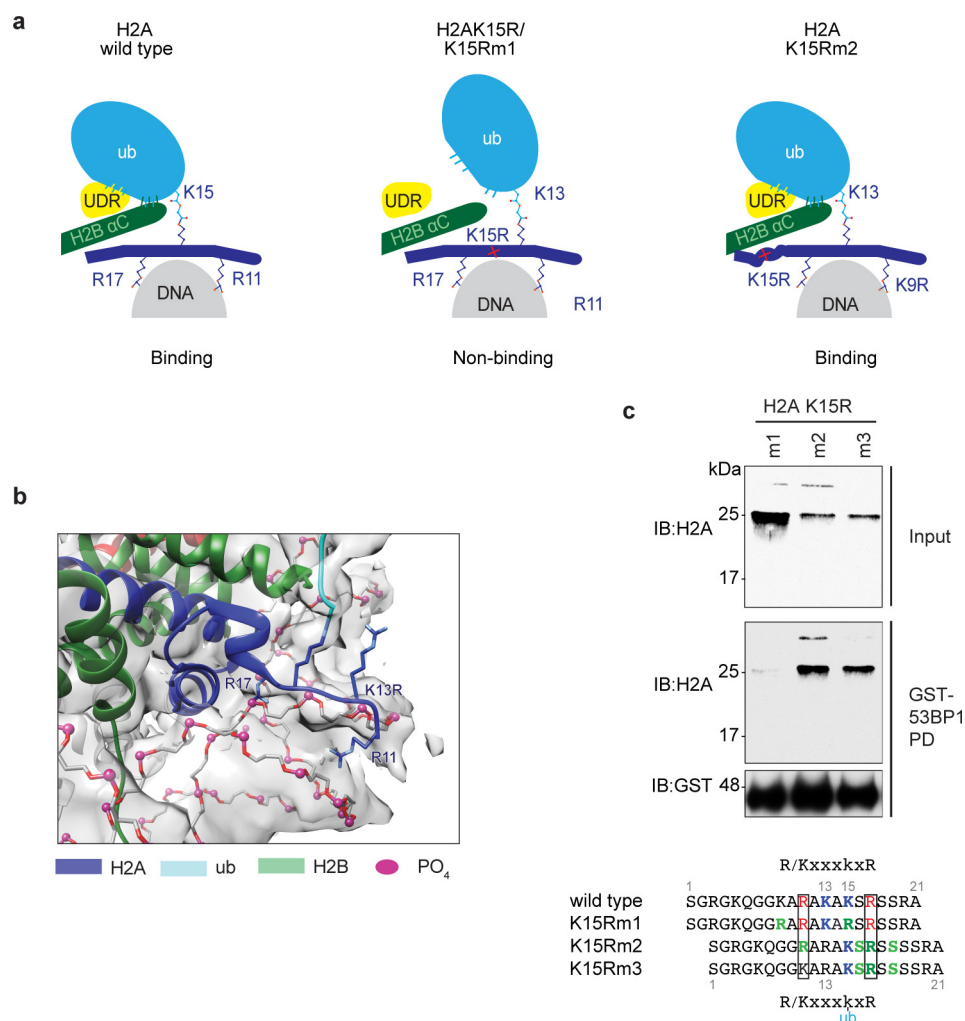


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Chemical ubiquitination of H2A and the constrained conformation of ubiquitin in the NCP-ubme-GST-53BP1 complex. **a**, Schematic of cross-linking reaction scheme between an electrophilic acetone (dibromoacetone, DBA) and two engineered cysteine residues in H2A and ubiquitin, respectively. TCEP was added to initially reduce disulfide bonds. **b**, Pilot reactions of cross-linkable ubiquitin mixed with H2A. Cross-linked products were separated by SDS-PAGE. H2A-only and ubiquitin-only reactions identify non-productive cross-linking in the final reaction, H2A-H2A and ub-ub. Correctly modified H2A is labelled H2AK_C15ub. Hexahistidine and TEV consensus sequence-tagged ubiquitin was used (HisTEV-ub) **c**, Chemically ubiquitinated H2AK_C15ub-containing NCPs interact with GST-53BP1. Immunoblot (IB) analysis of GST-53BP1 pull-down (PD) using NCPs assembled with unmodified H2A, catalytically ubiquitinated H2A or chemically ubiquitinated H2A (H2AK_C15ub). In this pilot experiment, the chemically ubiquitinated H2A runs with lower mobility due to the retention of the HisTEV tag. The tag was removed in all future experiments. **d**, Space-filling model of

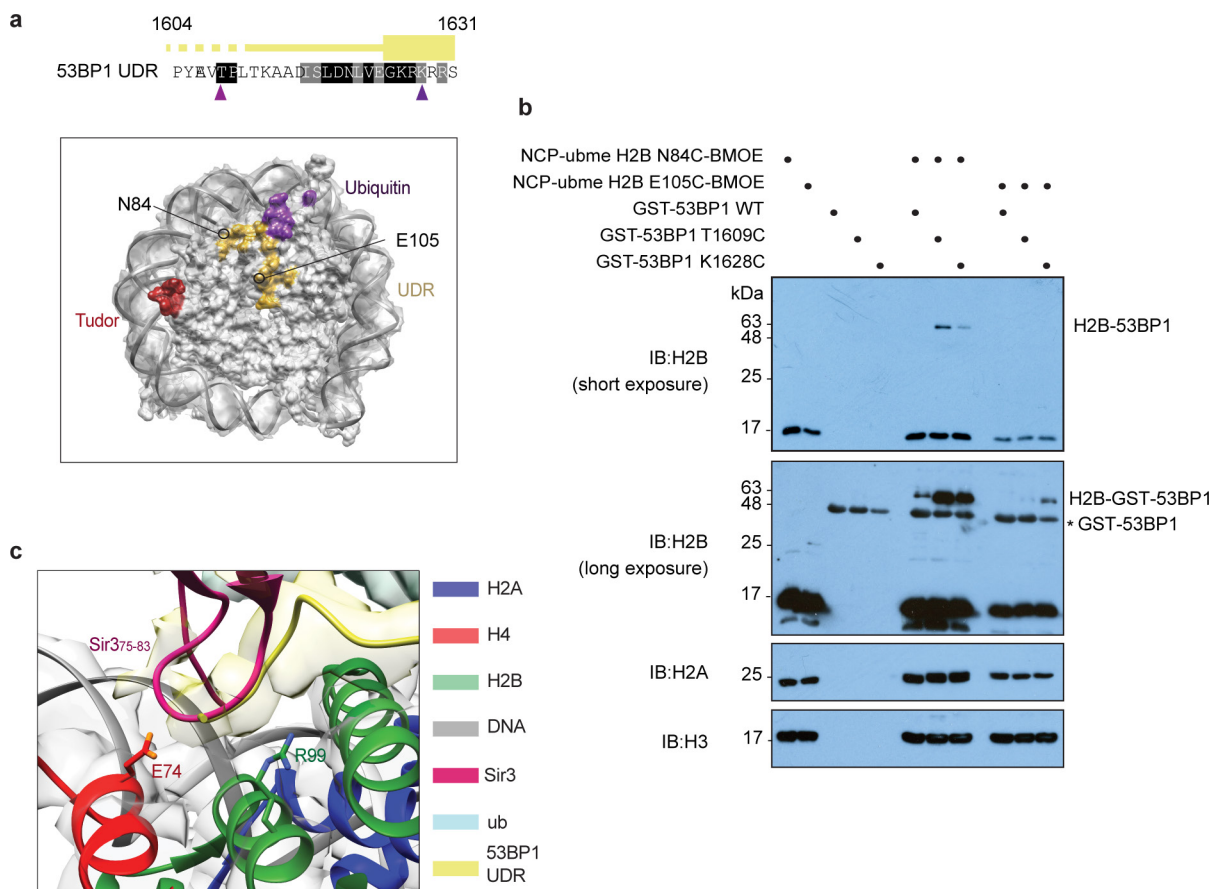
the covalently tethered ubiquitin, in a closed conformation, pulled over the surface of NCP. Key interacting histone residues are labelled.

e, Bio-layer interferometry traces of a single concentration of GST-53BP1 association and dissociation with immobilized NCP-ubme containing the indicated mutations in the H2B α C helix. Relative affinities are also shown. WT, wild-type H2B. **f**, Immunoblot analysis of GST-53BP1 pull-down assay to determine the effect of mutating the α C helix H2B residues that potentially form a hydrogen-bonding network with closed, 53BP1-bound ubiquitin. WT, wild-type H2B protein. **g**, Stained SDS-PAGE gel of purified reconstituted, nucleosomes used in this figure. H4K_C20me₂-modified NCPs containing cross-linked ubiquitin at indicated residues in H2A, H2B or both (left). These NCPs were used in assays in panels **e** and **f**. Biotinylated NCP-ubme complexes containing H2B variants used in the bio-layer interferometry assays (right). **h**, Immunoblot analysis of GST-53BP1 pull-down assay investigating the effect of H2BK120ub on GST-53BP1 binding to NCP-ubme.



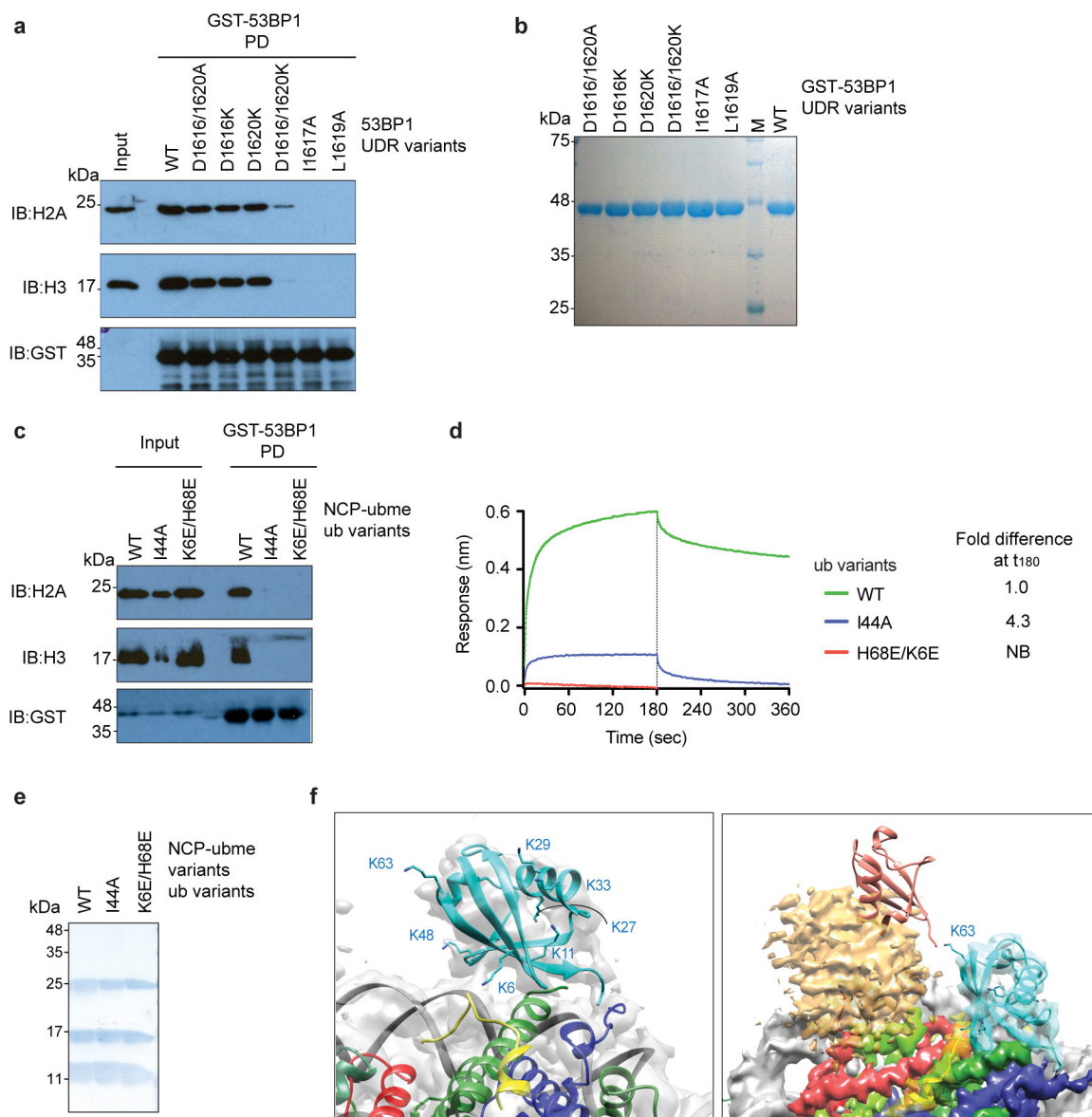
Extended Data Figure 6 | Structural basis of 53BP1 specificity for H2AK15ub. **a**, Diagrammatic representation of the arginine-fingers mechanism of 53BP1 recognition of H2AK15ub-containing NCPs. Sequences of H2A mutations are detailed in **c**. **b**, Top view of the H2A N-terminal tail, displaying the modelled arginines projecting into the nucleosomal DNA grooves. For clarity, only the DNA phosphodiester

backbone is shown. **c**, Immunoblot analysis of pull-down (PD) assay with RNF168-ubiquitinated NCPs containing the indicated H2A variants detailed at the bottom. GST-53BP1 can recognize H2AK13ub only when arginine 17 has been removed, allowing a shift in the N-terminal tail. Proposed R/KxxxKubxR consensus binding motif is indicated.



Extended Data Figure 7 | Specific orientation of the UDR region and H2B-H4 cleft interactions. **a**, Schematic of 53BP1 UDR region, indicating sites of engineered cysteines used for BMOE cross-linking (purple arrowheads) (top). Surface representation of modelled NCP with interaction interfaces coloured; tandem Tudor domain (H4 tail: red), UDR (H2B-H4 cleft, H2B α C helix and acidic patch: yellow) and ubiquitin (H2B α C helix and H2AK15: purple) (bottom). Locations of engineered H2B cysteine residues used for cross-linking are indicated (N84C and E105C). **b**, Immunoblot analysis of covalently cross-linked NCP-ubme-GST-53BP1 variants. BMOE, a bivalent maleimide cross-linker, was first reacted with NCP-ubme-containing H2B single cysteine variants, before

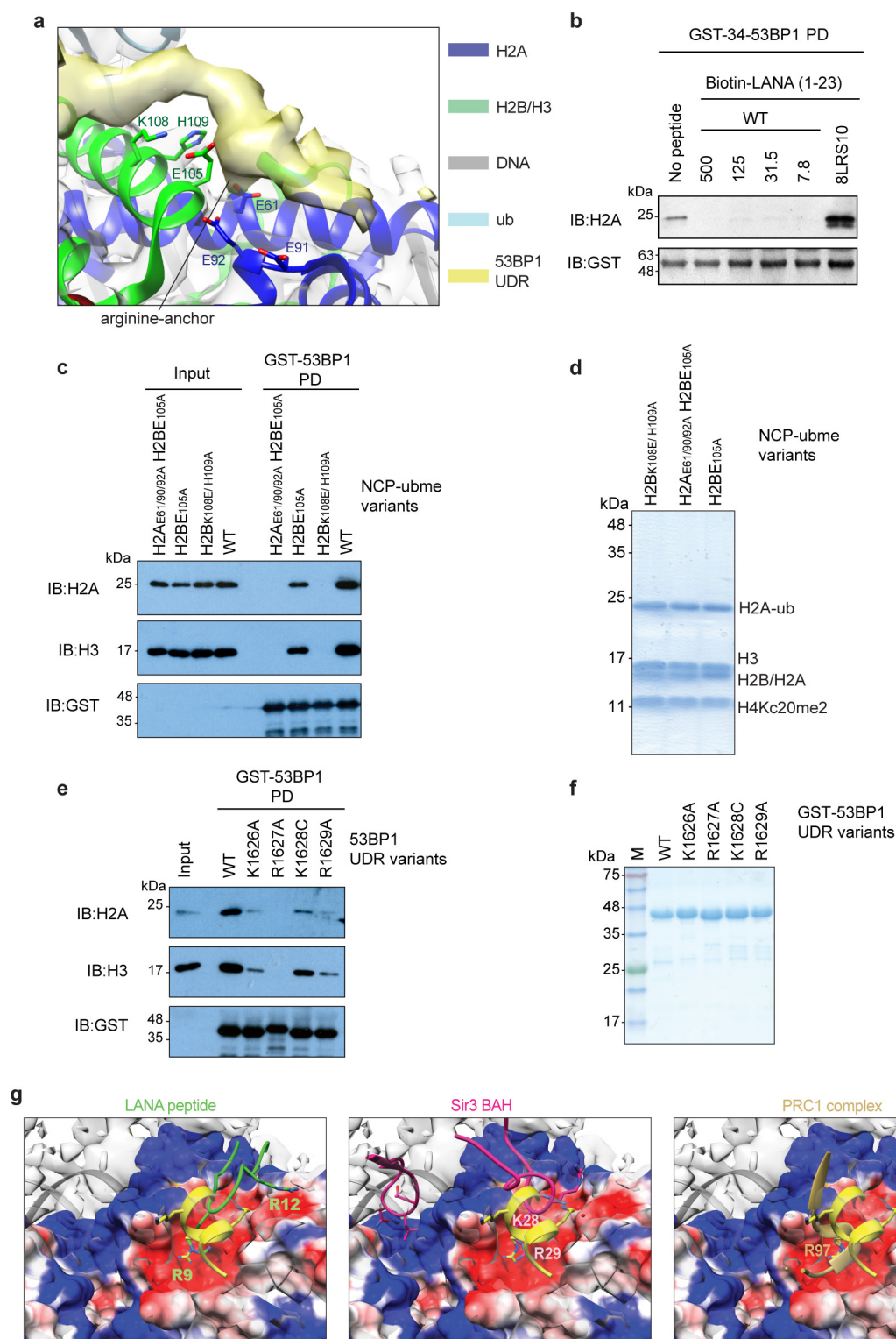
incubation with cysteine cross-linkable GST-53BP1 variants. Cross-linking to H2B is visualized by a shift in apparent molecular weight, equivalent to the addition of one GST-53BP1 moiety. The relatively weaker cross-linking of H2B(E105C) and GST-53BP1(K1628C) probably arises from the fact that the lysine is predicted to interact on the other face of the acidic patch. The asterisk denotes a non-specific band due to cross-reactivity of the anti-H2A antibody with non-cross-linked GST-53BP1. **c**, Magnified view of the H2B-H4 cleft at the rear of the NCP, with modelled UDR chain in the yellow density. Ribbon structure of Sir3 BAH domain (residues 75-83; Protein Data Bank accession code 3TU4)¹⁶, also proposed to interact in this region, was overlaid (purple).



Extended Data Figure 8 | Validation of the UDR-ubiquitin interaction.

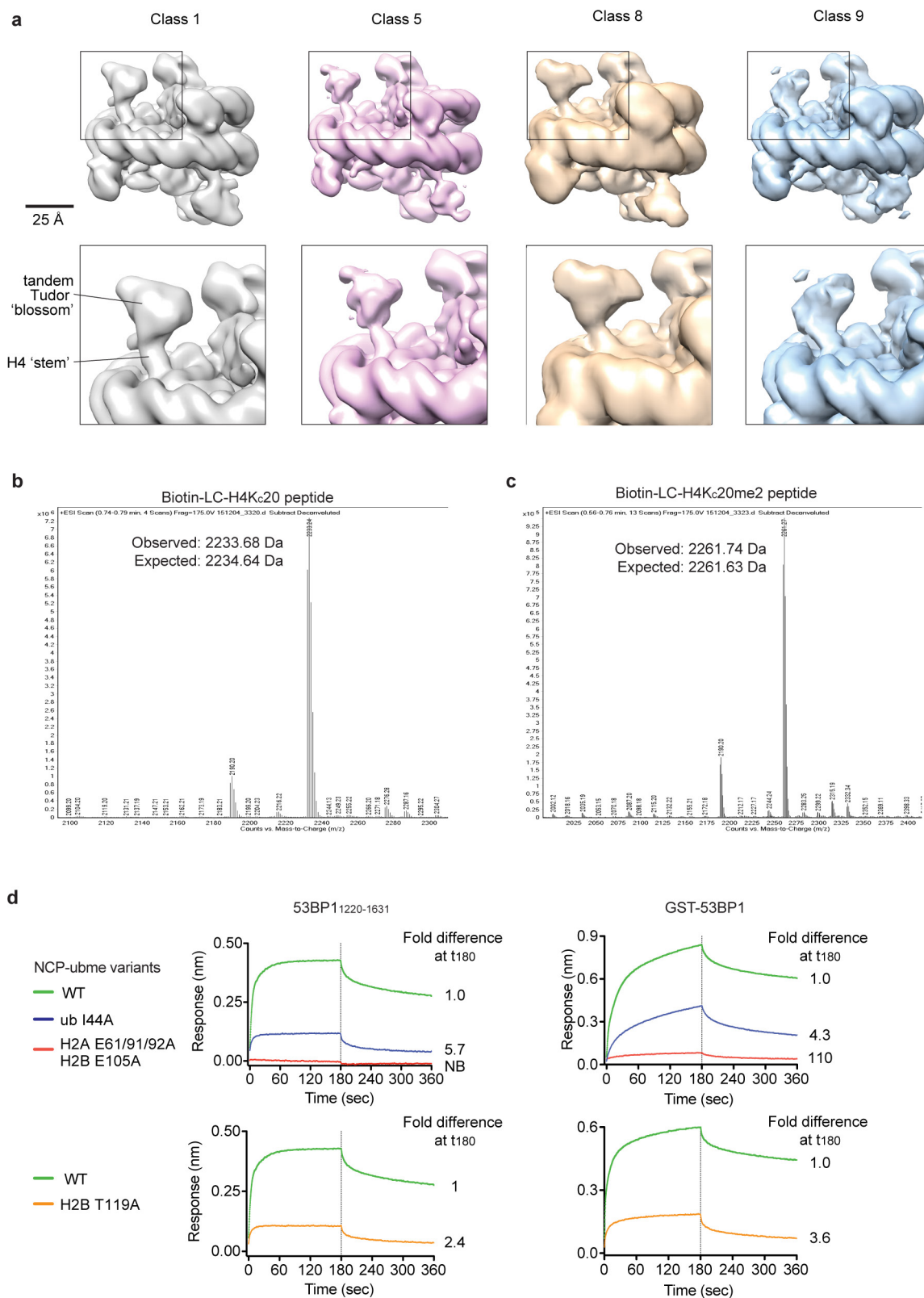
a, Immunoblot analysis of pull-down (PD) assays, immobilizing the indicated GST-53BP1 UDR mutations in residues 1616–1620 and monitoring NCP-ubme interaction. WT, wild-type GST-53BP1 protein. **b**, SDS-PAGE of purified GST-53BP1 UDR proteins used in the pull-down and bio-layer interferometry assays. M, molecular weight markers. **c**, Pull-down assays of GST-53BP1 with the indicated NCP-ubme variants. **d**, Bio-layer interferometry traces of NCP-ubme prepared with the indicated ubiquitin variants chemically ligated to H2A at position 15. NB, no binding detected. **e**, SDS-PAGE analyses of reconstituted, biotinylated nucleosomes used in **f**. **f**, Enlarged view of the 53BP1-bound constrained

modelled ubiquitin, with ball and stick representations of ubiquitin lysine residues indicated (left). Note the accessibility of Lys27 and Lys63, both reported to be following DSBs. Model of a Lys-3-linked di-ubiquitin (Protein Data bank accession code 3H7S)⁴⁸ built on H2AK15 within the NCPubme–GST-53BP1 structure (right). The distal ubiquitin (pink) projects away from the NCP surface towards the tandem Tudor domain of 53BP1, shown in orange. Although we note a minor steric clash between the modelled distal ubiquitin and the tandem Tudor domain, we surmise that the inherent flexibility of ubiquitin chains, coupled with the flexibility of the tethered Tudor domain on the H4 tail, probably enables 53BP1 to bind to NCPs with Lys63-linked ubiquitin chains on H2AK15.



Extended Data Figure 9 | 53BP1 UDR interactions with the nucleosome acidic patch. **a**, Magnified view of a ribbon representation of the NCP-ubme acidic patch with the overlaying density attributed to the UDR. **b**, Immunoblot analysis of a GST-34-53BP1 pull-down assay performed with RNF168-ubiquitinated NCP-me, in the presence of the acidic-patch-interacting LANA peptide (the GST-53BP1 protein used here has the 34 amino acid linker). The indicated amounts of LANA peptide were added as a competitor during the pull-down (concentration in μM). 8LRS10 peptide has negligible NCP binding¹⁹ and used as a control. **c**, Immunoblot analysis of GST-53BP1 pull-downs (PD) using NCP-ubme incorporating the indicated H2A and H2B mutants, which localize to the acidic patch and adjacent H2B αC helix. **d**, SDS-PAGE and InstantBlue staining to analyse purified reconstituted, biotinylated nucleosomes containing

H2A/H2B mutations used in the bio-layer interferometry assays. Compare to wild-type NCP-ubme in Extended Data Fig. 5g (right panel). **e**, Immunoblot analysis of GST-53BP1 pull-down assays, using selected 53BP1 UDR basic residue mutations. WT, wild-type GST-53BP1 protein. **f**, SDS-PAGE and InstantBlue staining to analyse purified GST-53BP1 UDR variant proteins used in the bio-layer interferometry assays in Fig. 3e. M, molecular weight markers. **g**, Enlarged view of UDR-acidic patch interaction site coloured according to coulombic surface charge, overlaid with the structure of other acidic patch chromatin binding factors: KSHV LANA peptide (Protein Data Bank accession code 1ZLA)¹⁹; the Sir3 BAH domain (Protein Data Bank accession code 3TU4)¹⁶ and the PRC1 complex (Protein Data Bank accession code 4R8P)¹⁵.



Extended Data Figure 10 | Flexibility of the 53BP1 tandem Tudor domain in the NCP-ubme structure and comparison of GST-53BP1 to 53BP1(1220–1631). **a**, A selection of aligned 3D maps obtained during determination of the NCP-ubme and GST-53BP1 structure, with an enlarged view of the density from the tandem Tudor domain of 53BP1, shown in the lower panels. Note that the position of the tandem Tudor domain density is highly variable, but is always tethered over the H4 N-terminal tail. **b**, 1D intact mass spectra of biotin-LC-H4₁₂₋₂₇ (K20C)

peptide chemically alkylated to create a lysine mimic. The reaction proceeded to near completion, but some unreacted peptide can be observed at 2,190 Da. **c**, 1D intact mass spectra of biotin-LC-H4₁₂₋₂₇ K20C, peptide chemically alkylated to create a dimethyl lysine mimic. The reaction proceeded to near completion, but some unreacted peptide can be observed at 2,190 Da. **d**, Bio-layer interferometry traces comparing the binding of GST-53BP1 with 53BP1(1220–1631) to NCP-ubme variants. Data from a single 53BP1 protein concentration is plotted.

Structure of the adenosine A_{2A} receptor bound to an engineered G protein

Byron Carpenter¹, Rony Nehmé¹, Tony Warne¹, Andrew G. W. Leslie¹ & Christopher G. Tate¹

G-protein-coupled receptors (GPCRs) are essential components of the signalling network throughout the body. To understand the molecular mechanism of G-protein-mediated signalling, solved structures of receptors in inactive conformations and in the active conformation coupled to a G protein are necessary^{1,2}. Here we present the structure of the adenosine A_{2A} receptor (A_{2A}R) bound to an engineered G protein, mini-G_s, at 3.4 Å resolution. Mini-G_s binds to A_{2A}R through an extensive interface (1,048 Å²) that is similar, but not identical, to the interface between G_s and the β₂-adrenergic receptor³. The transition of the receptor from an agonist-bound active-intermediate state^{4,5} to an active G-protein-bound state is characterized by a 14 Å shift of the cytoplasmic end of transmembrane helix 6 (H6) away from the receptor core, slight changes in the positions of the cytoplasmic ends of H5 and H7 and rotamer changes of the amino acid side chains Arg^{3.50}, Tyr^{5.58} and Tyr^{7.53}. There are no substantial differences in the extracellular half of the receptor around the ligand binding pocket. The A_{2A}R–mini-G_s structure highlights both the diversity and similarity in G-protein coupling to GPCRs⁶ and hints at the potential complexity of the molecular basis for G-protein specificity.

Structures of A_{2A}R bound to either inverse agonists^{7–9} or agonists^{4,5,10} have elucidated the molecular determinants of subtype specificity and ligand efficacy¹¹. However, the mechanism of activation of the receptor to allow G-protein coupling and the basis of G-protein selectivity is not fully understood. Structures of A_{2A}R in the inactive state have been determined bound to the antagonists ZM241385 (refs 7–9), XAC⁷, caffeine⁷ or 1,2,4-triazines¹², and all the structures are very similar. An intramembrane Na⁺ ion that can act as an allosteric antagonist was identified in the highest resolution structure (1.8 Å)¹³, and a homologous Na⁺ ion has been subsequently identified in other high-resolution structures of GPCRs^{14–16}. Four agonist-bound structures of A_{2A}R have also been determined after co-crystallization with either adenosine⁴, NECA⁴, CGS21680 (ref. 10) or UK432097 (ref. 5). All the structures are very similar and are thought to represent an active-intermediate conformation of the receptor, but not the fully active receptor that binds a G protein⁴. Observations that support this conclusion include the presence of rotamer changes of conserved amino acid residues associated with activation of other GPCRs¹⁷, and the absence of a large-scale movement of the cytoplasmic end of transmembrane helix 6 (H6) away from the receptor core¹¹. The G-protein-coupled state of A_{2A}R exhibits higher binding affinity for agonists compared to the uncoupled state¹⁸, but it is unclear whether the agonist-bound structures determined so far depict the binding pocket in a high-affinity or low-affinity conformation. Therefore, in order to elucidate the structure of the activated state of A_{2A}R, we have determined its structure bound to a high-affinity agonist and an engineered G protein.

There is a single reported structure of a GPCR bound to a heterotrimeric G protein, namely G_s-bound β₂-adrenergic receptor (β₂AR)³, which showed that virtually all the atomic contacts between the receptor and G protein were formed by the Gα subunit. To

facilitate the crystallization of any GPCR–G_s complex, we developed a minimal G protein, mini-G_s, that comprised a truncated form of the GTPase domain of Gα_s and included eight point mutations to stabilize the protein in the absence of Gβγ and in the presence of detergents (B.C. and C.G.T., manuscript submitted). In addition, three truncations removed the switch III region, 25 amino acids from the N terminus and the α-helical domain. Mini-G_s reproduced the

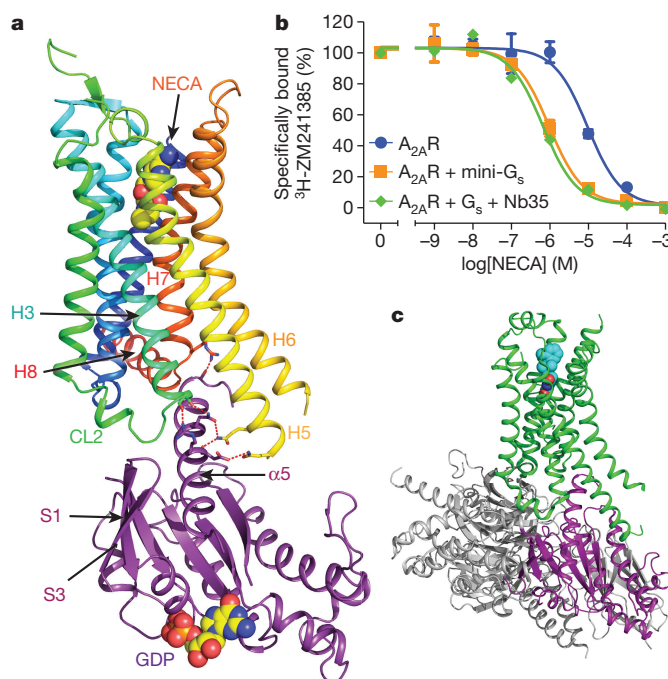


Figure 1 | Ligand binding and overall structure of the A_{2A}R–mini-G_s complex. **a**, The structure of A_{2A}R is depicted as a cartoon in rainbow coloration (N terminus in blue, C terminus in red) with mini-G_s in purple. The agonist NECA bound to A_{2A}R and GDP bound to mini-G_s are depicted as space-filling models (carbon, yellow; nitrogen, blue; oxygen, red; phosphorous, orange). Relevant secondary structural features are labelled. **b**, Mini-G_s increases the affinity of agonist binding to A_{2A}R similar to that observed by a heterotrimeric G protein. Competition binding curves were performed in duplicate ($n = 3$) by measuring the displacement of the inverse agonist ³H-ZM241385 by increasing concentrations of the agonist NECA (K_i values in parentheses, see Extended Data Fig. 1 for full data): blue circles, A_{2A}R (K_i 4.6 ± 0.3 μM); orange squares, A_{2A}R and mini-G_s (K_i 430 ± 80 nM); green diamonds, A_{2A}R and heterotrimeric G protein with nanobody Nb35 (K_i 340 ± 70 nM). Error bars represent s.e.m. G proteins were all added to membranes containing A_{2A}R to give a final concentration of 25 μM and the final concentration of NaCl was 100 mM (see Methods). **c**, The structure of β₂AR (green) bound to G_s (grey and purple) is depicted as a cartoon in the same orientation as A_{2A}R in **a**; the purple region in G_s corresponds to the structure of mini-G_s.

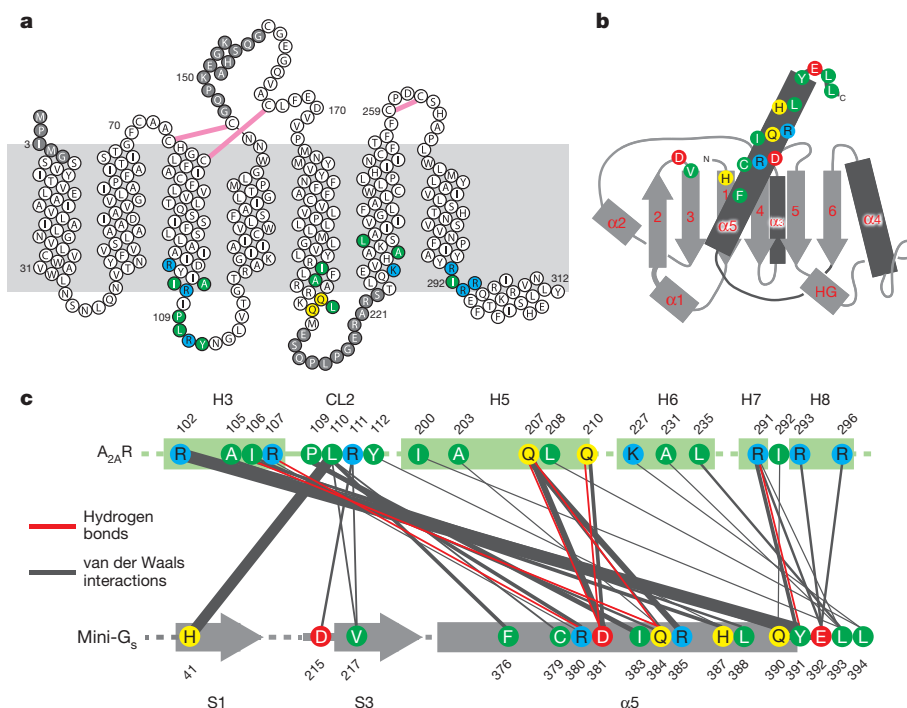


Figure 2 | Packing interactions between $A_{2A}R$ and mini- G_s . **a**, Diagram of $A_{2A}R$ depicting its secondary structure in the $A_{2A}R$ -mini- G_s structure. Residues shaded in grey are disordered in either chain A and/or chain B. Disulphide bonds are depicted as pink lines. **b**, Cartoon of the mini- G_s topology. **c**, Diagram of contacts between mini- G_s and $A_{2A}R$, with line thickness representing the relative number of interactions between amino acid residues. In all panels, amino acid residues depicted in colour are at the interface between mini- G_s and $A_{2A}R$ (within 3.9 Å), with colours reflecting the properties of the side chain; blue, positively charged; red, negatively charged; green, hydrophobic; yellow, hydrophilic.

increase in agonist affinity that occurred upon incubation of the receptor in the presence of the heterotrimeric G protein G_s and it also showed identical sensitivity to the presence of the allosteric antagonist Na^+

(Fig. 1 and Extended Data Fig. 1). In addition, mini- G_s readily formed a complex with $A_{2A}R$ in the presence of the agonist NECA and the complex was considerably more thermostable, particularly in

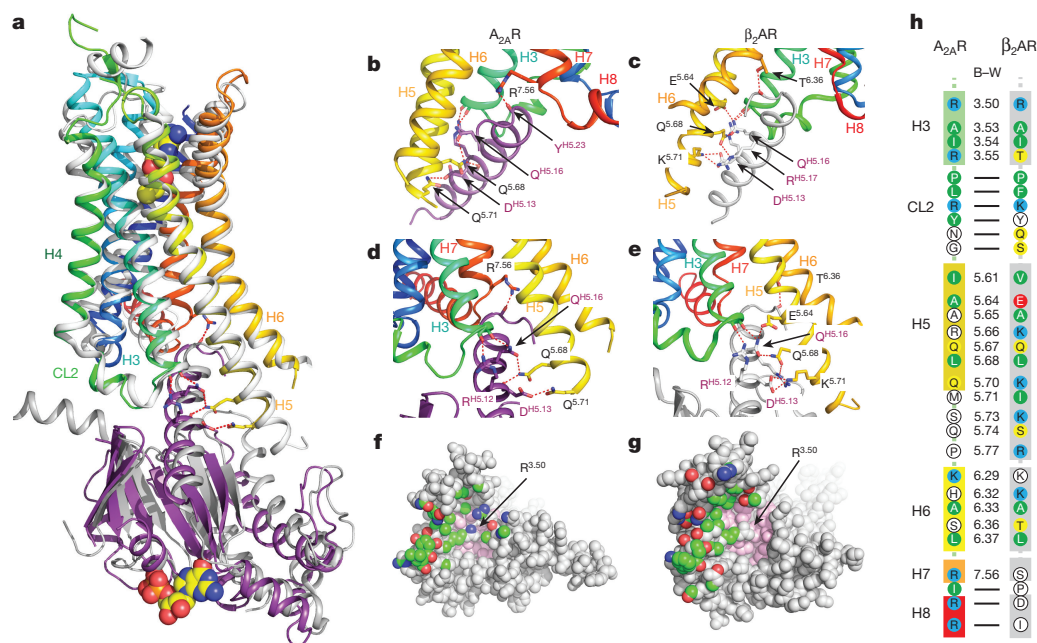


Figure 3 | Comparison of the $A_{2A}R$ -mini- G_s and β_2AR - G_s complexes. **a**, Structural alignment of β_2AR - G_s (PDB ID: 3SN6)³ and $A_{2A}R$ -mini- G_s was performed by aligning the receptors alone; $A_{2A}R$, rainbow colouration; β_2AR , grey. The resultant relative dispositions of $G\alpha_s$ (dark grey) bound to β_2AR and mini- G_s bound to $A_{2A}R$ (purple) are depicted. NECA and GDP are depicted as space-filling models (carbon, yellow; nitrogen, blue; oxygen, red; phosphorous, orange). The α -helical domains of $G\alpha_s$, $G\beta\gamma$ and Nb35 have all been omitted for clarity. **b–e**, Detailed comparisons of hydrogen bonds (red dashed lines) between the respective G proteins and receptors; both receptors are in rainbow colouration, with mini- G_s in purple and $G\alpha_s$ in grey. Labelling of amino acid residues shows the Ballesteros–Weinstein numbers for the receptors and the CGN notation for G proteins. **f, g**, Views of the cytoplasmic surface of $A_{2A}R$ and β_2AR , respectively, as space-filling models with atoms making contacts with their respective G proteins coloured according to their type; carbon, green; nitrogen, blue; oxygen, red. Atoms coloured pink comprise conserved hydrophobic residues in the core of the receptors against which Arg^{3.50} packs. **h**, Comparison of residues making contacts to G proteins in the $A_{2A}R$ -mini- G_s complex and the β_2AR - G_s complex. Amino acid residues in the receptors that make contacts are coloured: red, negatively charged; blue, positively charged; green, hydrophobic; yellow, hydrophilic. Residues in white are those that do not make contact to the respective G protein, but the equivalent residue in the other receptor does. Ballesteros–Weinstein (B–W) numbers are given for residues in transmembrane α -helices, with a dash for residues in loops or H8. Amino acid residues 5.71–5.77 are disordered in the $A_{2A}R$ -mini- G_s structure.

respectively, as space-filling models with atoms making contacts with their respective G proteins coloured according to their type; carbon, green; nitrogen, blue; oxygen, red. Atoms coloured pink comprise conserved hydrophobic residues in the core of the receptors against which Arg^{3.50} packs. **h**, Comparison of residues making contacts to G proteins in the $A_{2A}R$ -mini- G_s complex and the β_2AR - G_s complex. Amino acid residues in the receptors that make contacts are coloured: red, negatively charged; blue, positively charged; green, hydrophobic; yellow, hydrophilic. Residues in white are those that do not make contact to the respective G protein, but the equivalent residue in the other receptor does. Ballesteros–Weinstein (B–W) numbers are given for residues in transmembrane α -helices, with a dash for residues in loops or H8. Amino acid residues 5.71–5.77 are disordered in the $A_{2A}R$ -mini- G_s structure.

short-chain detergents, than A_{2A}R with only NECA bound (Extended Data Fig. 2). This complex was crystallized in the detergent octylthioglucoside by vapour diffusion, a data set collected from two crystals, the structure determined by molecular replacement (see Methods) and refined to 3.4 Å (Extended Data Table 1). Of the two A_{2A}R–mini-G_s complexes per crystallographic asymmetric unit, the density in complex AC was better defined and was therefore used for all subsequent analyses below (see Supplementary Discussion).

The A_{2A}R–mini-G_s complex contained density for the agonist NECA bound to A_{2A}R and density for a molecule of GDP bound to mini-G_s (Extended Data Fig. 3). The presence of GDP in the mini-G_s structure is a reflection of the properties of the engineered G protein, which, after complex formation, is insensitive to GTP-γS-mediated dissociation (B.C. and C.G.T., manuscript submitted). Mini-G_s is in a conformation virtually identical to that observed in the β₂AR–G_s structure (see below) and therefore represents an active state of the G protein, consistent with its ability to couple to A_{2A}R and induce high-affinity agonist binding (Fig. 1). The interface between A_{2A}R and mini-G_s is formed between 20 amino acid residues from the receptor and 17 residues in mini-G_s (Fig. 2, Extended Data Figs 4–6 and Supplementary Table 1), comprising a total buried surface area of 1,048 Å² on the receptor. In mini-G_s, contacts are made predominantly by the α5 helix involving 14 amino acid residues that pack against residues in H3, cytoplasmic loop 2 (CL2), H5, H6, H7 and H8 of A_{2A}R. Additional interactions include residues in S1, S3, the S2–S3 loop and α5 that form a hydrophobic pocket in which the side chain Leu110 in CL2 of A_{2A}R is sequestered (Extended Data Fig. 7). Amino acid residues in A_{2A}R and mini-G_s form complementary surfaces that pack together predominantly via van der Waals interactions (~90% of contacts) with six polar interactions across the interface. Helix α5 protrudes into the cleft within the cytoplasmic face of A_{2A}R created through the outward bending of the cytoplasmic end of H6. The apex of the α5 helix, Tyr391^{H5.23} (superscript refers to the CGN system for G proteins⁶) makes extensive van der Waals interactions with Arg102^{3.50} (superscript refers to the Ballesteros–Weinstein numbering system for GPCRs¹⁹) that forms the whole upper surface of the cleft (Fig. 3).

Superposition of the receptors in the A_{2A}R–mini-G_s complex and the β₂AR–G_s complex³ shows that the receptors have very similar architectures (r.m.s.d. 1.7 Å over 1,239 atoms). The intracellular faces of the receptors align very well, including the large outward shift of the cytoplasmic end of H6 on activation. However, mini-G_s does not superimpose exactly on the Gα subunit of the heterotrimeric G protein bound to β₂AR (Fig. 3). There is a difference in orientation of ~15°, although the difference is smaller (~10°) for the α5 helix. This is probably a consequence of the different amino acid residues in A_{2A}R compared to β₂AR (Fig. 3 and Extended Data Fig. 5), which results in a slightly different packing of the G proteins to the receptors, although we cannot discount the possible influence of lattice contacts. Alignment of mini-G_s with Gα_s bound to β₂AR shows that they are essentially identical (r.m.s.d. 0.92 Å over 1,158 atoms), with the most substantial difference being an 8° tilt between the respective α5 helices, resulting in a 3.7 Å displacement of the C_α of Tyr391 in mini-G_s away from the core of the G protein (Extended Data Fig. 8). Overall, there are 14 contacting residues in common between the β₂AR–G_s complex and the A_{2A}R–mini-G_s complex, with an additional 6 contacting residues present only in A_{2A}R and another ten present only in β₂AR (Supplementary Table 1). Many of the contacts between residues in the α5 helix of the G protein and the receptors are conserved, although the exact orientation and atomic contacts may differ (Fig. 3, Extended Data Fig. 6, Supplementary Table 1). Similarly, there is a highly conserved interaction between a hydrophobic residue in the centre of CL2, Leu110 in A_{2A}R and Phe139 in β₂AR, and residues His41^{S1.2}, Val217^{S3.1} and Asp215^{S2S3.1} in Gα_s (Extended Data Fig. 7). The main difference between the A_{2A}R–mini-G_s interface compared to the β₂AR–G_s interface occurs as a result of the different amino acid sequences at the H7–H8 boundary. In A_{2A}R, H7 terminates with Arg291^{7.56} and forms the sequence R^{7.56}IREFR (bold

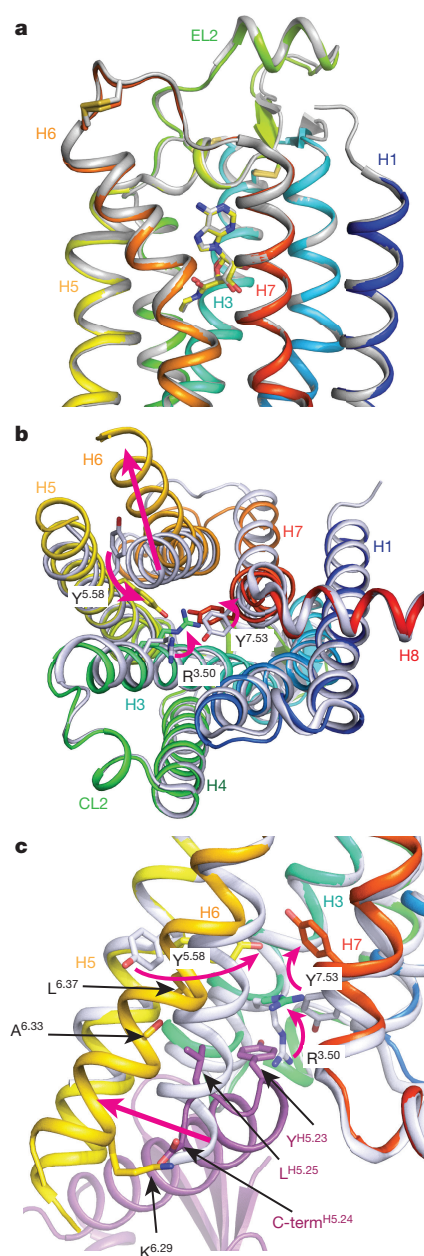


Figure 4 | Conformational changes in A_{2A}R upon G-protein binding.

A_{2A}R (rainbow colouration) bound to mini-G_s (purple) was aligned with A_{2A}R in the active-intermediate conformation bound to either NECA (PDB ID: 2YDV)⁴ or UK432097 (PDB ID: 3QAK)⁵ to highlight structural changes upon G-protein binding. Neither structure was used for both comparisons because the large extensions of the ligand UK432097 compared to NECA distorts the extracellular surface in comparison to the NECA-bound structure and the NECA-bound structure contains a thermostabilizing mutation in the intracellular half of the receptor. **a**, Alignment between 2YDV and the extracellular half of the A_{2A}R–mini-G_s complex is viewed parallel to the membrane plane. **b**, Alignment with 3QAK and viewed from the cytoplasmic surface with mini-G_s removed for clarity. **c**, Alignment with 3QAK viewed parallel to the membrane with the cytoplasmic side at the bottom. Residues are labelled with their Ballesteros–Weinstein numbers and arrows depict the direction of movement upon mini-G_s binding. Conversion of Ballesteros–Weinstein and CGN numbers to amino acid residues in A_{2A}R and mini-G_s, respectively, are as follows: R^{3.50}, Arg102; Y^{5.58}, Tyr197; K^{6.29}, Lys227; A^{6.33}, Ala231 carbonyl; L^{6.37}, Leu235; Y^{7.53}, Tyr288; Y^{H5.23}, Tyr391; L^{H5.25}, Leu393; C-term^{H5.26}, C terminus of mini-G_s (Leu394). The receptor is in rainbow colours and the mini-G_s is in purple.

amino acid residues make contact with mini-G_s), compared to the sequence S^{7,56}PDFRI in the equivalent position of β_2 AR, where none of the residues make contacts with G α_s . Another region of the receptors that differs in the presence/absence of contacts to their respective G proteins is at the end of H5, owing to the extension of H5 in β_2 AR by an additional turn compared to A_{2A}AR (Fig. 3, Extended Data Fig. 6 and Supplementary Table 1). From these examples it is clear that although the majority of amino acid residues at the interface between the receptor and G protein are identical, the specific atoms involved in the contacts differ either in terms of the amino acid side chains involved, their relative dispositions at the interface and/or the nature of the interaction.

Comparison of the active-intermediate state of UK432097-bound A_{2A}AR⁵ with the structure of A_{2A}AR bound to mini-G_s identified major rearrangements in the cytoplasmic half of the receptor core to accommodate G-protein binding (Fig. 4) and will be described in terms of the rearrangements required to transition from the active-intermediate state to the activated G-protein-bound conformation. First, the cytoplasmic end of H6 moves away from the receptor core by 14 Å as measured between the C α atoms of Thr224^{6,26} in the two different conformations. This movement is achieved through H6 bending outwards with little discernible rotation around the helix axis. The extent of H6 movement is dictated by van der Waals interactions between Lys227^{6,29}, Ala231^{6,33} and Leu235^{6,37} in A_{2A}AR and Leu393^{H5,25} and the carboxy terminus of mini-G_s. The movement of H6 requires substantial changes in the packing of the cytoplasmic end of H6 with helices H5 and H7. In particular, the side chains of highly conserved Tyr197^{5,58} and Tyr288^{7,53} both adopt new rotamers to fill the space previously occupied by the side chains of Leu235^{6,37} (the C α of which moves by 3.7 Å) and Ile238^{6,40} (C α moves by 2.2 Å), respectively. The shift in Tyr288^{7,53} allows Arg102^{3,50} of the conserved DRY motif to adopt a fully extended conformation, packing against the side chain of Tyr391^{H5,23} in the α 5 helix of mini-G_s. It is notable that the structural change from the inactive conformation to the active-intermediate state⁴ is characterized by the concerted rotation of H5, H6 and H7, whereas the conformation change from the active-intermediate state to the active conformation upon mini-G_s binding is characterized by the bending of H6 with little further rotation. In contrast to the considerable rearrangements of the cytoplasmic half of the receptor, there are no substantial changes in the extracellular half of the receptor (Fig. 4, Extended Data Fig. 9). Thus, the disposition of the ligand binding pocket described in the active-intermediate state most likely describes the high-affinity state of NECA bound to A_{2A}AR.

A_{2A}AR appears to have a very different energy landscape to the β -adrenergic receptors (β ARs). Both A_{2A}AR and β_2 AR exist in an ensemble of conformations whether bound to antagonists, agonists or to no ligand at all, and the presence of agonists increases the probability of formation of a fully active state^{20,21}. This active state is then stabilized by binding of a G protein^{3,22}. Structures of β ARs bound to agonists are all in a conformation very similar to the inactive state^{23,24}, whereas structures of A_{2A}AR bound to agonists are in an active-intermediate state^{4,5,10} very similar to the active state. Whether there is an active-intermediate state for β ARs equivalent to A_{2A}AR is unknown, but recently, it has been proposed based on extensive electron paramagnetic resonance spectroscopy data that β_2 AR also exists in two distinct states in the active conformation²⁰. This work shows that the active-intermediate and fully active states are distinct conformations in the intracellular half of the receptor. Given the highly conserved nature of the mechanism of GPCR activation, it is likely that the active-intermediate of A_{2A}AR may represent a common intermediate for many class A GPCRs, although it may exist only transiently depending on the energy landscape of the receptor.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 February; accepted 23 June 2016.

Published online 27 July 2016.

- Kobilka, B. K. Structural insights into adrenergic receptor function and pharmacology. *Trends Pharmacol. Sci.* **32**, 213–218 (2011).
- Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
- Lebon, G. *et al.* Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. *Nature* **474**, 521–525 (2011).
- Xu, F. *et al.* Structure of an agonist-bound human A_{2A} adenosine receptor. *Science* **332**, 322–327 (2011).
- Flock, T. *et al.* Universal allosteric mechanism for G α activation by GPCRs. *Nature* **524**, 173–179 (2015).
- Doré, A. S. *et al.* Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **19**, 1283–1293 (2011).
- Hino, T. *et al.* G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **482**, 237–240 (2012).
- Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
- Lebon, G., Edwards, P. C., Leslie, A. G. & Tate, C. G. Molecular determinants of CGS21680 binding to the human adenosine A_{2A} receptor. *Mol. Pharmacol.* **87**, 907–915 (2015).
- Lebon, G., Warne, T. & Tate, C. G. Agonist-bound structures of G protein-coupled receptors. *Curr. Opin. Struct. Biol.* **22**, 482–490 (2012).
- Congreve, M. *et al.* Discovery of 1,2,4-triazine derivatives as adenosine A_{2A} antagonists using structure based drug design. *J. Med. Chem.* **55**, 1898–1903 (2012).
- Liu, W. *et al.* Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **337**, 232–236 (2012).
- Miller-Gallacher, J. L. *et al.* The 2.1 Å resolution structure of cyanopindolol-bound β_1 -adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS One* **9**, e92727 (2014).
- Fenalti, G. *et al.* Molecular control of δ -opioid receptor signalling. *Nature* **506**, 191–196 (2014).
- Zhang, C. *et al.* High-resolution crystal structure of human protease-activated receptor 1. *Nature* **492**, 387–392 (2012).
- Huang, W. *et al.* Structural insights into μ -opioid receptor activation. *Nature* **524**, 315–321 (2015).
- Murphree, L. J., Marshall, M. A., Rieger, J. M., MacDonald, T. L. & Linden, J. Human A_{2A} adenosine receptors: high-affinity agonist binding to receptor-G protein complexes containing G β_3 . *Mol. Pharmacol.* **61**, 455–462 (2002).
- Ballesteros, J. A. & Weinstein, H. in *Methods in Neurosciences* Vol. 25 (eds Sealfon, S. C. & Conn, P. M.) 366–428 (Academic Press, 1995).
- Manglik, A. *et al.* Structural insights into the dynamic process of β_2 -adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).
- Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P. & Prosser, R. S. Activation of the A_{2A} adenosine G-protein-coupled receptor by conformational selection. *Nature* **533**, 265–268 (2016).
- Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469**, 175–180 (2011).
- Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist- β_2 adrenoceptor complex. *Nature* **469**, 236–240 (2011).
- Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β_1 -adrenergic receptor. *Nature* **469**, 241–244 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a grant from Heptares Therapeutics Ltd., the ERC (grant EMPSI 339995) and core funding from the Medical Research Council (MC_U105197215 and MC_U105184325). We thank the beamline staff at the European Synchrotron Radiation Facility (beamlines ID23-2, ID30-A3 and ID29) and at Diamond Light Source (beamline I24). We also thank R. Henderson, A. Jazayeri and T. Flock for comments on the manuscript.

Author Contributions B.C. performed receptor expression, purification, crystallization, cryo-cooling of the crystals, data collection, data processing and structure refinement. T.W. helped with crystallization, data collection and data processing. R.N. performed the stability assays and pharmacological analyses on A_{2A}AR–mini-G_s complexes. A.G.W.L. was involved in data processing and structure solution, refinement and analysis. Manuscript preparation was performed by B.C., A.G.W.L. and C.G.T. The overall project management was by C.G.T.

Author Information Atomic co-ordinates and structure factors for the A_{2A}AR–mini-G_s complex have been submitted to the Protein Data Bank (PDB) under accession code 5G53. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.G.T. (cgt@mrc-lmb.cam.ac.uk).

Reviewer Information Nature thanks O. Ernst, K. Jacobson and K.-N. Klotz for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Expression and purification of mini-G_s. The mini-G_s construct used (construct 414) was identical to mini-G_s (construct 393) described elsewhere (B.C. and C.G.T., manuscript submitted), except that one additional mutation, L63Y, was included to improve crystal quality. An N-terminal histidine tag (His₆) and TEV protease cleavage site were present to facilitate purification. Mini-G_s was expressed in *E. coli* strain BL21(DE3)RIL upon induction with IPTG (50 µM) for 20 h at 25 °C. Cells were harvested by centrifugation and lysed by sonication in lysis buffer (40 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 10 mM imidazole, 5 mM MgCl₂, 50 µM GDP, 1 mM PMSF, 2.5 µM Pepstatin-A, 10 µM Leupeptin, 50 µg/ml DNase I, 100 µg/ml lysozyme, 100 µM DTT), supplemented with Complete protease inhibitors (Roche). The lysate was clarified by centrifugation and loaded onto a 10 ml Ni²⁺ Sepharose FF column. The column was washed with wash buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 40 mM imidazole, 10% glycerol, 1 mM MgCl₂, 50 µM GDP) and eluted with elution buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM MgCl₂, 50 µM GDP). TEV protease was added and the sample was dialysed overnight against dialysis buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 1 mM MgCl₂, 10 µM GDP). TEV protease was removed by negative purification on Ni²⁺-NTA resin (Qiagen). The sample was concentrated to 1.5 ml and loaded onto a Superdex-200 26/600 gel-filtration column, equilibrated with gel-filtration buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 1 mM MgCl₂, 1 µM GDP, 0.1 mM TCEP). Peak fractions were pooled and concentrated to 100 mg/ml. The pure protein was aliquoted, flash-frozen in liquid nitrogen, and stored at -80 °C. A typical yield was 100 mg of pure mini-G_s per litre of culture.

Expression and purification of adenosine A_{2A}R. Wild-type human adenosine A_{2A}R (residues 1–308) was modified to contain a C-terminal histidine tag (His₁₀) and TEV protease cleavage site. The N154A mutation was introduced to remove a potential N-linked glycosylation site. Recombinant baculoviruses expressing A_{2A}R were prepared using the flashBAC ULTRA system (Oxford Expression Technologies). *Trichoplusia ni* cells were grown in suspension in ESF921 media (Expression Systems) to a density of 3 × 10⁶ cells/ml, infected with A_{2A}R baculovirus and incubated for 72 h. Cells were harvested and membranes prepared by two ultracentrifugation steps in 20 mM HEPES pH 7.5, 1 mM EDTA, 1 mM PMSF, NECA (100 µM), NaCl (300 mM), PMSF (1 mM) and Complete protease inhibitors (Roche) were added to the membranes, and the sample was mixed for 30 min at room temperature. Membranes from 3 l of cells were solubilized with 2% *n*-decyl-β-D-maltopyranoside (DM) on ice for 1 h. The sample was clarified by ultracentrifugation and loaded onto a 5 ml Ni-NTA column (Qiagen). The column was washed with wash buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 80 mM imidazole, 100 µM NECA, 0.15% DM), and eluted with elution buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 300 mM imidazole, 100 µM NECA, 0.15% DM). The eluate was concentrated using a 50 kDa cut-off Amicon centrifugal ultrafiltration unit (Millipore), and exchanged into desalting buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol, 100 µM NECA, 0.15% DM) using a PD10 column (GE Healthcare). TEV protease was added, and the sample was incubated on ice overnight. TEV protease was removed by negative purification on Ni²⁺-NTA resin. The sample was concentrated to 0.2 ml and loaded onto a Superdex 200 column (GE Healthcare). Peak fractions were pooled and concentrated to approximately 20 mg/ml. A typical yield was 2 mg of pure A_{2A}R per litre of culture.

Complexation and crystallization. Purified A_{2A}R was mixed with a 1.2-fold molar excess of mini-G_s. MgCl₂ (1 mM) and apyrase (0.1 U) were added, and the mixture was incubated on ice overnight. The sample was diluted tenfold in gel-filtration buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 100 µM NECA, 0.35% *n*-octyl-β-D-thioglucopyranoside OTG), concentrated to 0.2 ml, and loaded on to a Superdex 200 column (pre-equilibrated in the same buffer). Peak fractions, containing the A_{2A}R–mini-G_s complex, were pooled and concentrated to 20 mg/ml. The A_{2A}R–mini-G_s complex was crystallized by vapour diffusion in OTG either in the presence or absence of cholesterol hemisuccinate (CHS), but there was no discernible difference in the quality of crystals that grew under the two different conditions (the structure was determined using data collected from two crystals, one from each condition). Crystallization plates were set up at 4 °C using 120-ml

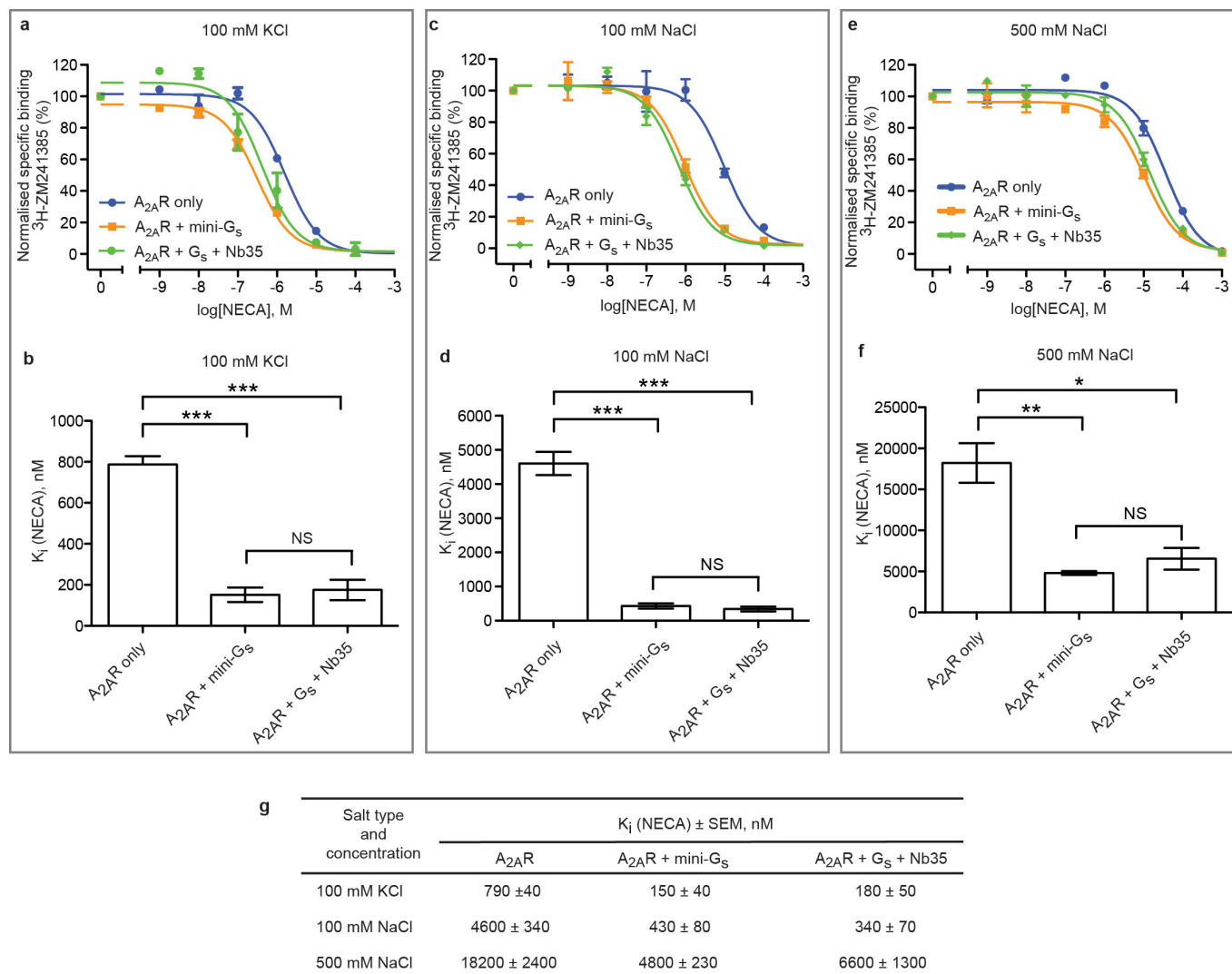
sitting drops. Crystals used for structure solution were grown in two conditions, either: 0.1 M NaOAc pH 5.5, 10% PEG 2000 (in the presence of CHS); or 0.1 M NaOAc pH 5.7, 9.5% PEG 2000 MME (in the absence of CHS). Crystals were cryo-protected in mother liquor supplemented with 30% PEG 400 and flash frozen in liquid nitrogen.

Data collection, processing and refinement. Diffraction data were collected at the European Synchrotron Radiation Facility on beamline ID23-2 with a Pilatus 2M detector, using a 6 µm × 8 µm microfocus beam (0.8729 Å wavelength). Data were collected using either standard or helical collection modes. Data from two crystals were used for structure solution. Data were processed using MOSFLM²⁵ and AIMLESS²⁶. The structure was solved by molecular replacement with PHASER²⁷ using the structures of the thermostabilized A_{2A}R (PDB ID: 2YDV)⁴ and the Gα_s GTPase domain (residues 40–59 and 205–394) from the β₂AR–G_s complex (PDB ID: 3SN6)³ as search models. Model refinement and rebuilding were performed using REFMAC²⁸ and COOT²⁹.

Competition binding assay. FreeStyle HEK293-F cells transiently expressing wild type A_{2A}R were resuspended in either assay buffer A (25 mM HEPES, pH 7.5, 100 mM KCl, 1 mM MgCl₂), assay buffer B (25 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM MgCl₂), or assay buffer C (25 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM MgCl₂), and were lysed by 10 passages through a 26-gauge needle. Purified binding partners were buffer-exchanged to the respective buffer before being added to the membranes at a final concentration of 25 µM. The mixture was aliquoted and NECA was added (0 to 1 mM final concentration, prepared in assay buffers containing 1 U/ml apyrase). The samples were incubated for 90 min at 22 °C. ³H-ZM241385 was added at its apparent K_d (2.5 nM) and allowed to bind for a further 90 min at 22 °C. Non-specific binding was determined in the presence of 100 µM of ZM241385. Receptor-bound and free radioligand were separated by filtration through 96-well GF/B filter plates (pre-soaked with 0.1% polyethyleneimine), and washed three times with the appropriate buffer. Plates were dried and radioactivity was quantified by liquid scintillation counting using a Tri-Carb 2910 TR (Perkin Elmer). Data were analysed by nonlinear regression using GraphPad Prism software. The K_i for NECA binding was derived from one-site fit K_i analysis. Data from at least three independent experiments, each performed in duplicate, were analysed using an unpaired two-tailed *t*-test for statistical significance.

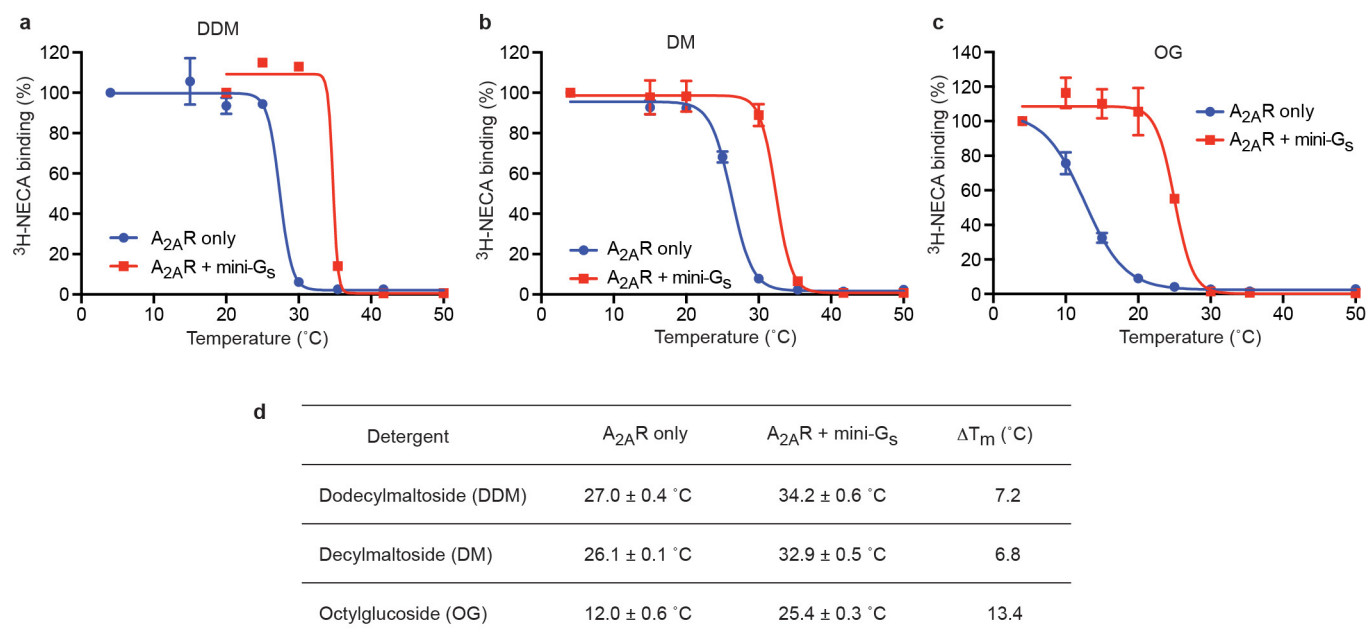
Thermostability assay. Membranes from *Trichoplusia ni* cells expressing wild-type human A_{2A}R were resuspended in T_m buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 1 mM MgCl₂) and homogenized by ten passages through a 26-gauge needle. Binding partner was added at a final concentration of 25 µM. ³H-NECA and unlabelled NECA were mixed in a molar ratio of 1:5 and added to the membranes to give a final concentration of 1 µM (approximately tenfold above the apparent K_d). The samples were incubated at room temperature for 1 h, then chilled on ice for 30 min. DDM, DM or OG were added to a final concentration of 0.1%, 0.13% or 0.8%, respectively, and samples were incubated on ice for 1 h. Cell debris and insoluble material were removed by centrifugation for 5 min at 20,000g and the supernatant was aliquoted into PCR strips. Samples were heated to the desired temperature for exactly 30 min, then quenched on ice for 30 min. Samples (50 µl) were loaded onto gel-filtration resin packed into a 96-well filter plate (Millipore), which was centrifuged to separate receptor-bound from free radioligand. Non-specific binding was determined in the presence of 200 µM unlabelled NECA. Radioactivity was quantified by liquid scintillation counting using a MicroBeta TriLux scintillation counter (PerkinElmer). Data were analysed by nonlinear regression using GraphPad Prism software. Apparent T_m values were derived from sigmoidal dose–response analysis. Results represent the mean ± s.e.m. of two independent experiments, performed in duplicate.

25. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
26. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
27. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
28. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
29. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
30. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60**, 2288–2294 (2004).



Extended Data Figure 1 | Pharmacological analyses of $\text{A}_{2\text{A}}\text{R}$ -mini-G_s complexes. Competition assays were performed on $\text{A}_{2\text{A}}\text{R}$ expressed in HEK293 cell membranes with the agonist NECA competing for the binding of radiolabelled inverse agonist $^3\text{H-ZM241385}$. Experiments were performed in the presence of either 100 mM KCl (**a**, **b**), 100 mM NaCl (**c**, **d**) or 500 mM NaCl (**e**, **f**) to confirm the similar behaviour of mini-G_s

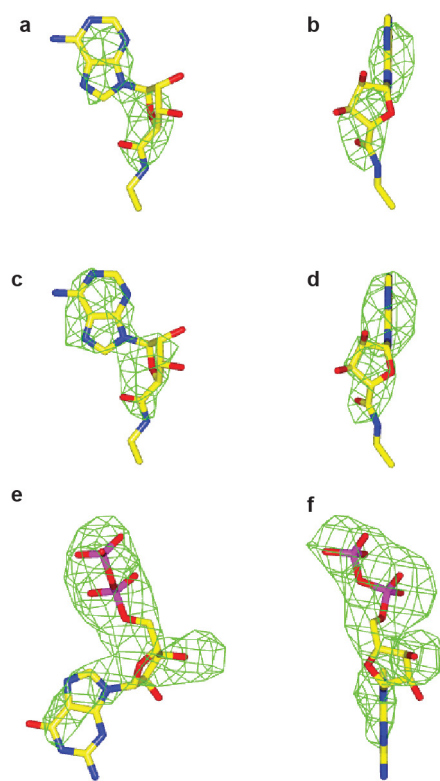
with heterotrimeric G_s with nanobody Nb35 for stabilization of the complex. Results are summarized in the Table (**g**). All error bars represent the s.e.m. for at least three independent experiments performed in duplicate. Comparisons of data in **b**, **d** and **f** were performed using an unpaired *t*-test with significance denoted by asterisks: ****P* < 0.0001; ***P* < 0.01; **P* < 0.1; not significant (NS) *P* > 0.1.



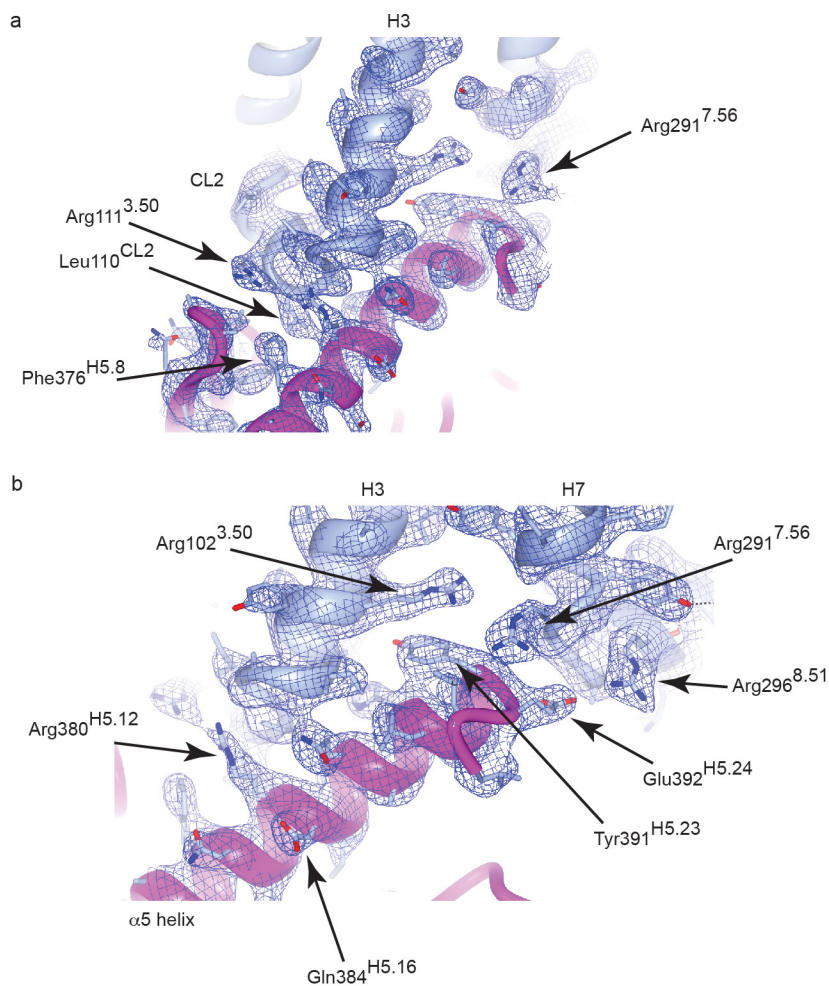
Extended Data Figure 2 | Thermostability of detergent-solubilized ³H-NECA-bound A_{2A}R in the presence or absence of mini-Gs.

Unpurified A_{2A}R was solubilized in detergent at the following concentrations: **a**, DDM, 0.1%; **b**, DM, 0.13%; **c**, OG, 0.8%. Samples were heated for 30 min, quenched on ice and the amount of ³H-NECA bound

determined. Data were analysed by nonlinear regression and apparent T_m values (transition temperature where 50% of the receptor is inactive) were determined from analysis of the sigmoidal dose–response curves fitted (**d**). Results represent the mean ± s.e.m. of two independent experiments, performed in duplicate.

**Extended Data Figure 3 | Omit maps for NECA and GDP.**

a–f, Orthogonal views of omit map difference density for NECA in $A_{2A}R$ chain A (**a**, **b**), NECA in $A_{2A}R$ chain B (**c**, **d**) and GDP in mini- G_s chain C (**e**, **f**). The contour level is 2.5σ in panels **a–d** and 3.0σ in panels **e** and **f**. Figures were made using CCP4mg³⁰.



Extended Data Figure 4 | Electron density for the interface region of the A_{2A}R–mini-G_s complex. The backbones of A_{2A}R and mini-G_s are shown in cartoon representation in light blue and magenta respectively. Side chains are shown in stick representation (carbon, light blue; oxygen, red; nitrogen, deep blue). The electron density of the final $2F_o - F_c$ map

is shown contoured at 1.2σ . For clarity, transmembrane helices H5 and H6 and the corresponding electron density have been omitted. **a**, View showing the interaction between the C-terminal helix of mini-G_s and the CL2 loop of A_{2A}R. **b**, View showing the interactions between side chains of the C-terminal helix of mini-G_s and three Arg residues of A_{2A}R.

```

      10      20      30      40      50
sp|P04896  MGCLGNSKTEDQRNEEKAQREANKKIEKQLQKDKQVYRATHRLLLLLGAGE
Mini Gs C  -----GIEKQLQKDKQVYRATHRLLLLLGADN
Mini Gs D  -----GIEKQLQKDKQVYRATHRLLLLLGADN
                *****.:
890123456789012345678901234567890123  1234567
HN2      3      4      5      |      S1      |

      60      70      80      90     100
sp|P04896  SGKSTIVKQMRILHVNGFNGEGGEEDPQAARSNSDGEKATKVQDIKNNLK
Mini Gs C  SGKSTIVKQMRIYHGGSGGSGG-----
Mini Gs D  SGKSTIVKQMRIYHGGSGGSGG-----
                ***** * . ...*
                1234567
                H1      |

      110     120     130     140     150
sp|P04896  EAIETIVAAMSNLVPVELANPENQFRVDYILSVMNVPDFDFPPEFYEHA
Mini Gs C  -----
Mini Gs D  -----

      160     170     180     190     200
sp|P04896  KALWEDEGVRACYERSNEYQLIDCAQYFLDKIDVIKQDDYVPSDQDLLRC
Mini Gs C  -----
Mini Gs D  -----
89012 123456789012  1234567890123  123456
HC1 | HD      |      HE      1      |      HF      |

      210     220     230     240     250
sp|P04896  RVLTSGLFETKFQVDKVNFMFMDVGGQDERRKWIQCFNDVTAIIFVVAS
Mini Gs C  ---TSGIFETKFQVDKVNFMFMDVGGQDERRKWIQCFNDVTAIIFVVD
Mini Gs D  ---TSGIFETKFQVDKVNFMFMDVGGQDERRKWIQCFNDVTAIIFVVD
                ***** *
                12345678 12345678 1234567890 1234567
                S2      | S3      | H2      | S4      |

      260     270     280     290     300
sp|P04896  SSYNMVIREDNQTNRLQEALNLFKSIWNNRWLRTISVILFLNKQDLLAEK
Mini Gs C  SDY-----NRLQEALNDFKSIWNNRWLRTISVILFLNKQDLLAEK
Mini Gs D  SDY-----NRLQEALNDFKSIWNNRWLRTISVILFLNKQDLLAEK
                *.*
                *****
                123456789012345678 1234567 1234567
                H3      1      | S5      | HG

      310     320     330     340     350
sp|P04896  VLAGKSKIEDYFPEFARYTTTPEDATPEPGEDPRVTRAKYFIRDEFRLRIST
Mini Gs C  VLAGKSKIEDYFPEFARYTTTPEDATPEPGEDPRVTRAKYFIRDEFRLRIST
Mini Gs D  VLAGKSKIEDYFPEFARYTTTPEDATPEPGEDPRVTRAKYFIRDEFRLRIST
                *****
8901234567 123456789012345678901234567
HG 1      |      H4      1      2      |

      360     370     380     390 394
sp|P04896  ASGDGRHYCYPHFTCAVDTENIRRVFNDCRDI IQRMHLRQYELL
Mini Gs C  ASGDGRHYCYPHFTCAVDTENARRIFNDCRDI IQRMHLRQYELL
Mini Gs D  ASGDGRHYCYPHFTCAVDTENARRIFNDCRDI IQRMHLRQYELL
                *****
                12345 12345678901234567890123456
                S6      H5      1      2      2

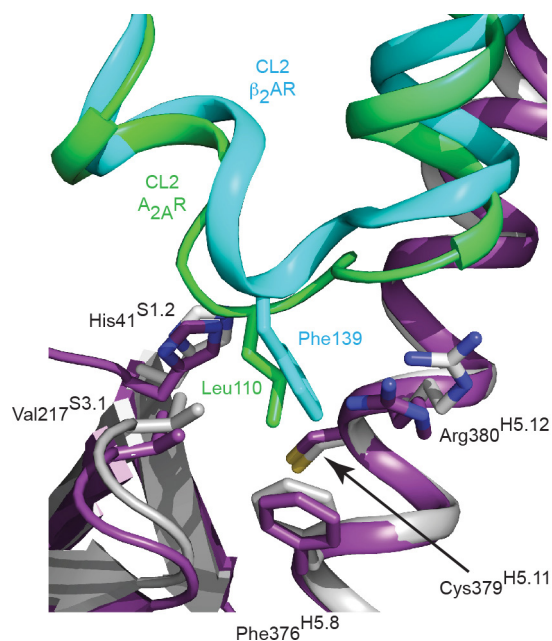
```

Extended Data Figure 5 | Alignment of mini-G_s with GNAS2. Comparison of amino acid residues in mini-G_s (chains C and D) within 3.9 Å of A_{2A}R (green) in the A_{2A}R–mini-G_s structure and the amino acid residues in bovine GNAS2 (P04896) within 3.9 Å of β₂AR in the β₂AR–G_s structure (turquoise). The CGN system is used for reference.

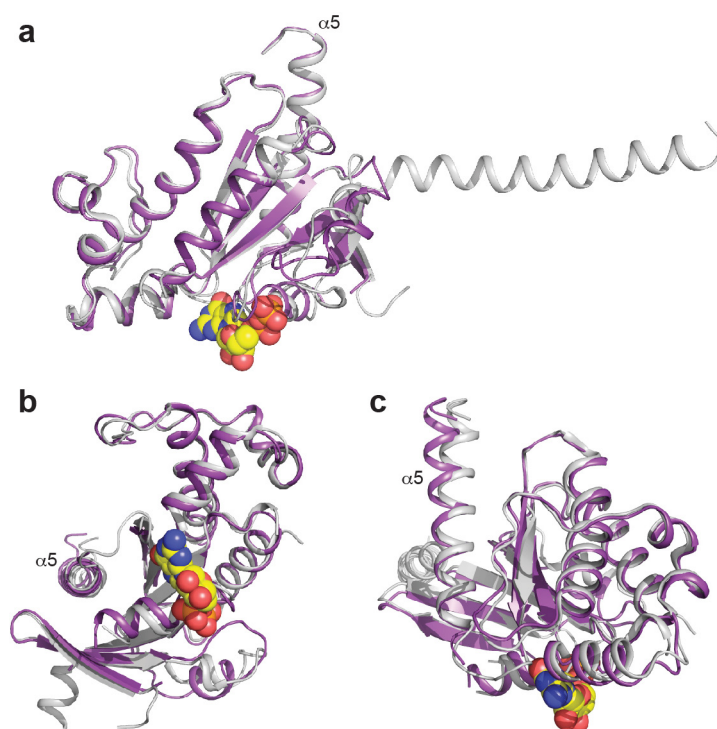
				1.50			
adrb2_human	1	MGQPGNGSAFLLAPNRSHAPDHDVTQQRDE	VWVGMGIVMSLIVLAI	VFGN	NVLVITAI	AK	60
AA2AR_human	1	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWL					33
AA2AR_human A	1	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWL					33
AA2AR_human B	1	MPIMGSSVYITVELAIAVLAILGNVLVCWAVWL					33
		* * . . ***** *					
		2.50					
adrb2_human	61	FERLQTVTNYFITS	LACADL	VLMGLAVVPFGAAHILMKM	WTFGNFWCEF	WTSIDVLCVTAS	120
AA2AR_human	34	NSNLQNV	TNYFV	VS	LAAADI	AVGVLAIPFAITISTGF--CAACHGCLFIACFVLVLTQSS	91
AA2AR_chain A	34	NSNLQNV	TNYFV	VS	LAAADI	AVGVLAIPFAITISTGF--CAACHGCLFIACFVLVLTQSS	91
AA2AR_chain B	34	NSNLQNV	TNYFV	VS	LAAADI	AVGVLAIPFAITISTGF--CAACHGCLFIACFVLVLTQSS	91
		** .***** .***.***. .*. .**.					
		3.50		4.50			
adrb2_human	121	IETLCVIAVD	RYFAITS	PFFKYQS	LLTKNKARV	IILMVIVSGLTSFLPIQMHWRATHQE	180
AA2AR_human	92	IFSL	LAI	AIDRYIAIRI	PLRYNGLVTGTRAKGIIAICWVLSFAIGLTPMLGWNNCGQPKE	151	
AA2AR_chain A	92	IFSL	LAI	AIDRYIAIRI	PLRYNGLVTGTRAKGIIAICWVLSFAIGLTPMLGWNNCGQPKE	151	
AA2AR_chain B	92	IFSL	LAI	AIDRYIAIRI	PLRYNGLVTGTRAKGIIAICWVLSFAIGLTPMLGWNNCGQPKE	151	
		* .* .***.***. ** * .*. *. * .*. * .*. * *					
		5.50					
adrb2_human	181	AIN-----	CYANETCCDFFTNQAYAI	ASSIVSFYVPLVIMVFVYSRV	VFQ	EAKRQLQKI	233
AA2AR_human	152	GKNHSQ	CGEGQVACLFEDV	VP	MNYMVYFNFFACVLVPLLLMLGVYLRIFLAARQLKQM	211	
AA2AR_chain A	152	GK	AHSQCGEGQVACLFEDV	VP	MNYMVYFNFFACVLVPLLLMLGVYLRIFLAARQLKQM	211	
AA2AR_chain B	152	GK	AHSQCGEGQVACLFEDV	VP	MNYMVYFNFFACVLVPLLLMLGVYLRIFLAARQLKQM	211	
		* * . . * *					
		. ***.***. ** * * .***					
		6.50					
adrb2_human	234	DKSEGR	RFHVQ	NLSQVEQDGR	TGHGLRRSSKFC	LKEHKALKKTLGIIMGTFTLCWL	PFFIVN
AA2AR_human	212	ESQPLPGERARS	-----			TLQKEVHAAKSLAIIVGLFALCWL	PLHIIN
AA2AR_chain A	212	ESQPLPGERARS	-----			TLQKEVHAAKSLAIIVGLFALCWL	PLHIIN
AA2AR_chain B	212	ESQPLPGERARS	-----			TLQKEVHAAKSLAIIVGLFALCWL	PLHIIN
		** .* *. * * * * .***** *					
		7.50					
adrb2_human	294	IVHVIQDNLI--	RKEVYILLNWIGYVNSGFN	PLIYCRSPDFRIAFQELLCLR	SSLKAYG	351	
AA2AR_human	254	CFTFFCPDCSHAPLWLMYLAIVLSHTNSV	VN	NPIYAYRIREFRQ	TRFKIIRSHVLRQ	QEP	
AA2AR_chain A	254	CFTFFCPDCSHAPLWLMYLAIVLSHTNSV	VN	NPIYAYRIREFRQ	TRFKIIRSHVLE	NLYF	
AA2AR_chain B	254	CFTFFCPDCSHAPLWLMYLAIVLSHTNSV	VN	NPIYAYRIREFRQ	TRFKIIRSHVLE	NLYF	
		. . . * * * *					
adrb2_human	352	NGYSSNGNTGEQSGYHVEQEKENKLLCEDLPGTEDFVGHQGTVP	SDNIDSQGRNCSTNDS	411			
AA2AR_human	314	FKAAGTSARVLA	AHGS	DGEQVSLRLNGHPPGVWANGSAPHERRP	NGYALGLVSGGSAQE	373	
AA2AR_chain A	314	QGH	HHHHHHHHHHH	325			
AA2AR_chain B	314	QGH	HHHHHHHHHHH	325			
adrb2_human	412	LL	413				
AA2AR_human	374	SQGNTGLPDVELLSHELKGVCP	PEPPGLDDPLAODGAGVS	412			

Extended Data Figure 6 | Alignment of β_2 AR and A_{2A}R amino acid sequences. adrb2_human, human β_2 -adrenergic receptor; AA2AR_human, human adenosine A_{2A} receptor; AA2AR chain A, chain A of the crystallized A_{2A}R–mini-G_s structure; AA2AR chain B, chain B of the crystallized A_{2A}R–mini-G_s structure. Residues in the receptors that are within 3.9 Å of either G α in the β_2 AR–G_s complex or mini-G_s in the A_{2A}R–mini-G_s complex are highlighted in turquoise or green, respectively.

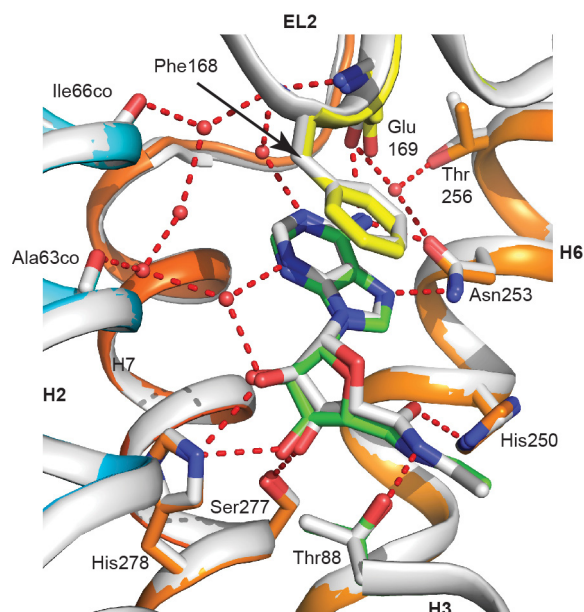
Key Ballesteros–Weinstein numbers are shown in blue and mutations in the crystallized A₂A R to facilitate purification and crystallization are shown in red. Grey bars indicate the positions of α -helices in the β_2 AR–G_s structure, whereas red bars represent these regions in the A₂A R–mini-G_s structure; where there is a discrepancy in helix length between chain A and B of A₂A R, the bar is coloured pink.



Extended Data Figure 7 | A conserved hydrophobic binding pocket at the receptor- $G\alpha_s$ interface. The A_{2A} R-mini- G_s complex was aligned to the β_2 AR- G_s complex via the receptors; A_{2A} R, green; β_2 AR, turquoise; mini- G_s (purple); $G\alpha_s$ (grey).



Extended Data Figure 8 | Comparison between receptor-bound mini-G_s and Gα_s. a–c, Three different views of an alignment of mini-G_s (chain C, purple) bound to A_{2A}R with the GTPase domain of Gα_s (grey) bound to β₂AR. GDP bound to mini-G_s is depicted as a space filling model (carbon, yellow; oxygen, red; nitrogen, blue; phosphorus, orange). The α5 helix that interacts with the receptors is labelled.



Extended Data Figure 9 | Comparison of the NECA binding site in the active-intermediate state compared to the mini- G_s -bound state. The structure of NECA-bound $A_{2A}R$ (grey cartoon, with the carbon atoms of NECA also in grey) in the active-intermediate state was aligned with the structure of the NECA-bound $A_{2A}R$ -mini- G_s complex (rainbow colouration, with the carbon atoms of NECA in green). Key amino acid residues for both receptors are depicted (sticks; carbon atoms in the same colour as the respective receptor) that form hydrogen bonds (red dashed line) with either NECA or the associated water network (red spheres). Note that the water molecules depicted are from only the NECA-bound $A_{2A}R$ structure in the active-intermediate state, because the resolution of the $A_{2A}R$ -mini- G_s structure was insufficient to identify water molecules. Carbonyl oxygens are denoted by 'co' after the residue name.

Extended Data Table 1 | Data collection and refinement statistics

Data collection	
Space group	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions <i>a</i> , <i>b</i> , <i>c</i> (Å)	90.6, 111.8, 161.3
Resolution (Å) ¹	40.3-3.4 (3.49-3.40)
<i>R</i> _{merge}	0.173 (0.747)
<i>I</i> / σ <i>I</i>	3.6 (1.2)
Completeness (%)	90.6 (78.5)
Redundancy	2.6 (2.4)
Refinement	
Resolution (Å)	40.3-3.4
No. reflections	19788
<i>R</i> _{work} / <i>R</i> _{free} (%)	28.4/31.5
No. atoms	7359
Protein	7248
Ligand/detergent/nucleotide	44/40/27
Water	0
B-factors (Å ²)	
Protein	79.9
Ligand/detergent/nucleotide	67.9/98.6/69.0
R.M.S.D.	
Bond lengths (Å)	0.008
Bond angles (°)	1.15

¹Values in parentheses are for the highest resolution shell.

Architecture of fully occupied GluA2 AMPA receptor–TARP complex elucidated by cryo–EM

Yan Zhao^{1*}, Shanshuang Chen^{1*}, Craig Yoshioka², Isabelle Baconguis¹ & Eric Gouaux^{1,3}

Fast excitatory neurotransmission in the mammalian central nervous system is largely carried out by AMPA-sensitive ionotropic glutamate receptors¹. Localized within the postsynaptic density of glutamatergic spines, AMPA receptors are composed of heterotetrameric receptor assemblies associated with auxiliary subunits, the most common of which are transmembrane AMPA receptor regulatory proteins (TARPs). The association of TARPs with AMPA receptors modulates receptor trafficking and the kinetics of receptor gating and pharmacology². Here we report the cryo-electron microscopy (cryo-EM) structure of the homomeric rat GluA2 AMPA receptor saturated with TARP $\gamma 2$ subunits, which shows how the TARPs are arranged with four-fold symmetry around the ion channel domain and make extensive interactions with the M1, M2 and M4 transmembrane helices. Poised like partially opened ‘hands’ underneath the two-fold symmetric ligand-binding domain (LBD) ‘clamshells’, one pair of TARPs is juxtaposed near the LBD dimer interface, whereas the other pair is near the LBD dimer–dimer interface. The extracellular ‘domains’ of TARP are positioned to not only modulate LBD clamshell closure, but also affect conformational rearrangements of the LBD layer associated with receptor activation and desensitization, while the TARP transmembrane domains buttress the ion channel pore.

Stargazin is the founding member of the TARPs³, a family of membrane proteins related in amino acid sequence to claudin, a four-helix transmembrane protein⁴. Co-expression of recombinant AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors with TARPs largely recapitulates native receptor gating kinetics, ion channel properties, and pharmacology, consistent with the notion that TARPs are fundamental components of neuronal AMPA receptor signalling complexes⁵, yet with a heterogeneous stoichiometry ranging from one to four TARPs per receptor⁶. Stargazin, also known as TARP $\gamma 2$, modulates AMPA receptor gating by slowing deactivation and desensitization, accelerating the recovery from desensitization, increasing the efficacy of partial agonists such as kainate, and attenuating polyamine block of calcium-permeable AMPA receptors^{7–9}. Despite progress in visualization of the AMPA receptor–TARP complex at a low resolution¹⁰, determination of the molecular architecture of the AMPA receptor–TARP $\gamma 2$ complex and defining a molecular mechanism for TARP modulation of receptor function have proven elusive, in part because TARPs are bound weakly to the receptor and dissociate under typical conditions employed in complex solubilization and purification.

X-ray crystal and single-particle cryo-EM structures of AMPA receptors show that they are tetrameric assemblies consisting of three layers: the amino-terminal domain (ATD), the LBD and the transmembrane domain (TMD)^{11–15}. Whereas the ATDs and LBDs assemble as two-fold symmetric dimers-of-dimers^{16,17}, the TMDs adopt four-fold symmetry, which results in a symmetry mismatch between the TMD and the LBD and gives rise to two-fold related conformationally distinct subunit pairs, A–C and B–D¹¹. Each LBD resembles a ‘clamshell’¹⁸ that

is open in apo and antagonist-bound states and closes upon binding of agonists¹⁹. Structures of the GluA2 receptor in agonist-bound pre-open states illustrate that the LBDs are assembled in a ‘back-to-back’ fashion, with agonist-induced closure of the LBDs causing a separation of the LBD–TMD linkers and a translation of the LBD layer closer to the membrane^{12,13}. The agonist-bound desensitized state, by contrast, undergoes a massive rearrangement of the ATD and LBD layers, thus decoupling agonist binding from ion channel gating^{13,14,20}.

To define the molecular basis for TARP modulation of AMPA receptor gating and pharmacology, we sought to elucidate the architecture of the AMPA–TARP $\gamma 2$ complex by single-particle cryo-EM. Here we focus on the wild-type homomeric rat GluA2 AMPA receptor²¹, bearing an arginine at the Q/R site²² and harbouring the flop splice variant²³, where we have co-expressed the receptor in mammalian cells in combination with full-length TARP $\gamma 2$ (ref. 24, 25). Evidence for formation of a physiologically relevant receptor–TARP $\gamma 2$ complex in these cells was shown by an increase in the efficacy of the partial agonist kainate to $80 \pm 2\%$ of that of a full agonist, glutamate²⁶ (Fig. 1a). To define conditions for solubilization and purification of AMPA receptor fully bound with TARPs, we carried out fluorescence-detection size-exclusion chromatography (FSEC)²⁷ studies on mammalian cells co-expressing GluA2 receptor and an engineered enhanced green fluorescent protein (eGFP)–TARP $\gamma 2$ fusion. By systematic screening of detergents and lipids through FSEC, we found that whereas dodecyl maltopyranoside (DDM) leads to dissociation of the receptor–TARP $\gamma 2$ complex, digitonin retains the complex integrity, allowing TARP to remain associated with receptor following solubilization and purification (Extended Data Fig. 1a). We proceeded to purify the native GluA2 receptor and full-length TARP complex in the presence of the competitive antagonist MPQX²⁸ (Extended Data Fig. 1b, c), succeeding in isolating a homogeneous population suitable for single-particle cryo-EM analysis (Extended Data Fig. 1d, e).

Three-dimensional (3D) reconstruction of the receptor–TARP $\gamma 2$ complex without the imposition of symmetry revealed an overall architecture consistent with previous crystal and cryo-EM structures of the antagonist-bound GluA2 receptor^{11,14} (Fig. 1b). The initial 3D classification yielded four classes, one of which had four protrusions on the extracellular side of the detergent micelle, related by an approximate four-fold axis of symmetry, and was composed of the largest number of particles. The remaining three classes had poorly resolved features associated with the extracellular domains and did not exhibit four-fold symmetric protrusions from the micelle, features associated with the presence of TARP subunits, and thus were excluded from the analysis (Extended Data Fig. 2). Further studies, and larger data sets, will be required to elucidate the structures of additional structural classes of the receptor–TARP $\gamma 2$ complex.

To improve the density of the TARPs and the structural features of receptor–TARP interactions, we carried out focused refinement of the LBD and TMD layers, masking the conformationally heterogeneous

¹Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239, USA. ²Department of Biomedical Engineering, Oregon Health & Science University, 2730 SW Moody Ave., Portland, Oregon 97201, USA. ³Howard Hughes Medical Institute, Oregon Health & Science University, Portland, Oregon 97239, USA.

*These authors contributed equally to this work.

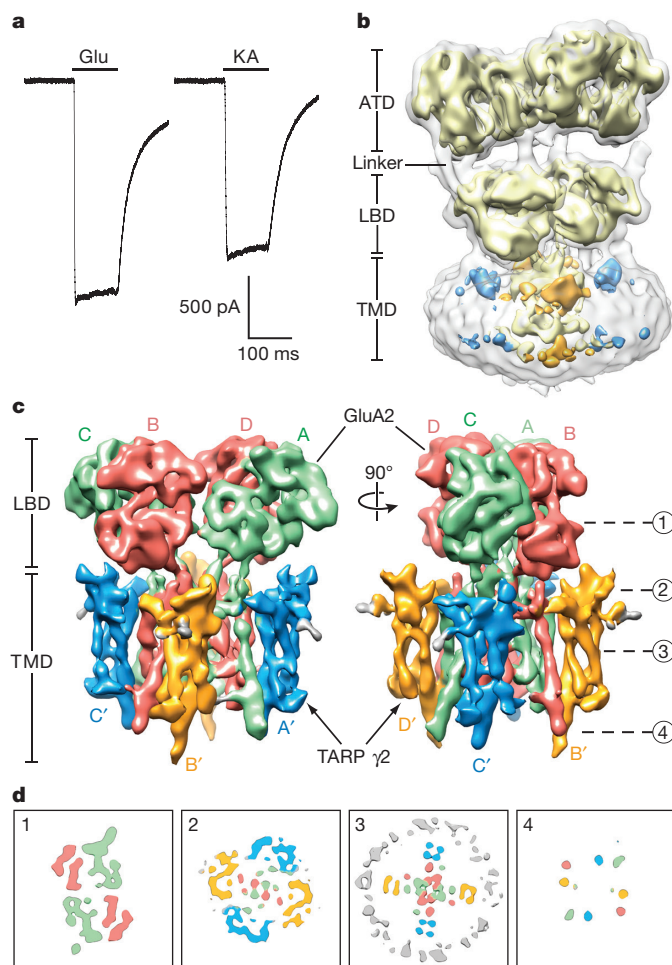


Figure 1 | Function and reconstruction of GluA2-TARP $\gamma 2$ complex. **a**, Whole-cell patch clamp recordings from cells expressing the GluA2-TARP $\gamma 2$ complex. A representative pair of currents recorded using the same cell is shown. The ratio between steady-state currents evoked by kainate and glutamate is 0.80 ± 0.02 (mean \pm s.d., $n = 5$). **b**, Initial 3D reconstruction of GluA2-TARP $\gamma 2$ complex contoured at lower (outer) and higher (inner) threshold levels showing distinct features for the receptor and associated TARP. **c**, Refined 3D reconstruction focused on LBD and TMD layers, where the ATDs were excluded from refinement. A-C and B-D subunits of the GluA2 receptor are in green and red, respectively; TARP $\gamma 2$ associated with receptor A-C and B-D subunits are in blue and gold, respectively. **d**, Cross-sections of the EM map at LBD layer, TARP-LBD interface layer, TMD layer and C-terminal layer at indicated 'height', with density features coloured as in panel **c**.

ATD layer, with application of C2 symmetry coincident with the two-fold axis that relates the LBD dimers and the four-fold axis of the TMD, in the subsequent 3D reconstructions and refinements (Extended Data Fig. 2). The resulting density map has an estimated resolution of 7.3 Å (Extended Data Fig. 3) and illustrates hallmark features of the LBD clamshells and the receptor TMD. Most importantly, the density map clearly reveals the presence of four TARPs, arranged with four-fold symmetry, surrounding the exterior of the receptor TMD, consistent with a fully saturated receptor-TARP complex (Fig. 1c). The receptor-TARP $\gamma 2$ complex features a similar symmetry mismatch between the two-fold related LBD layer and four-fold related TMD layer as found in isolated receptor (Fig. 1d), with TARP subunit pairs A'-C' and B'-D' 'underneath' the A-C and B-D LBDs, respectively. We further note that density for the full-length receptor M4 and carboxy-terminal TARP transmembrane (TM) helices extends into the cytoplasm (Fig. 1d).

To generate the structural model of GluA2 receptor-TARP $\gamma 2$ complex, we extracted individual LBDs and the intact TMD from the MPQX-bound GluA2 crystal structure¹¹ and fit them into the cryo-EM

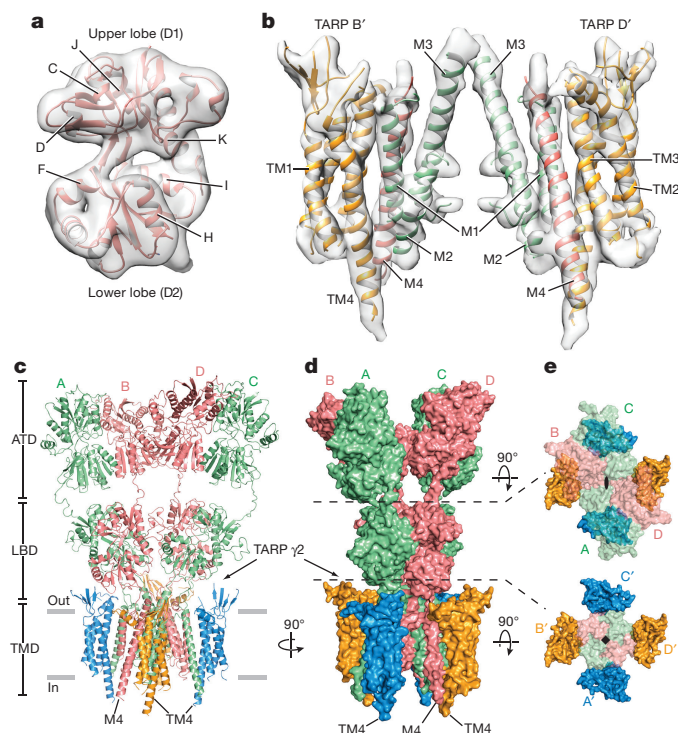


Figure 2 | Structure of GluA2 receptor in complex with TARP $\gamma 2$. **a**, **b**, Cryo-EM density maps for LBD (**a**) and TM (**b**) regions of the complex. A'-C' TARPs and associated receptor TMs were omitted for clarity. **c**, **d**, Ribbon diagram (**c**) and surface representation (**d**) of the complex. The carboxy termini of selected TARP helices (TM4) and receptor helices (M4) are labelled. **e**, Two 'top-down views', in which elements of the structure above the indicated dashed lines have been omitted, for clarity, from the figures. The GluA2 receptor is shown in transparent surface representation to allow visualization of the TARP subunits.

density as rigid bodies (Fig. 2a, b). Manual adjustments of secondary structure elements were applied where there was supporting density, followed by fitting of the pre-M1 and M3-S2 linkers (Fig. 2c, d). The S2-M4 linkers were not visible in the density maps. The cryo-EM density for the ATDs was poorly resolved, in line with their flexibility. Thus, we did not focus on optimizing the density of the ATD layer, concentrating instead on the crucial LBD, TMD and TARP regions. The degree of LBD clamshell opening is similar to the MPQX-bound full-length receptor structure¹¹, confirming that the complex is stabilized in an antagonist-bound state (Fig. 2a).

Because *de novo* structure determination for TARP was not feasible at the resolution of this study, we generated a homology model of the TARP $\gamma 2$ using the claudin-19 crystal structure as a template⁴ (Extended Data Fig. 4a). Rigid body fitting of the TARP model in the density map was unambiguous, driven by strong helical density for the TARP TMs and consistent with computational analysis of the TARP density (Extended Data Fig. 4b-d). The particularly long TM4 of TARP, which protrudes into the cytoplasm, provides an additional structural landmark to validate the fitting of the TARP model to the experimental density map. Whereas there was strong density to support the presence of a TARP β -sheet on the extracellular side of the membrane, like that found in claudin-19 (ref. 4) (Extended Data Fig. 5), little density was found for the loop connecting the $\beta 1$ and $\beta 2$ strand, or the loop between TM3 and TM4 (Extended Data Fig. 4a). These loops were therefore excluded from the homology model (Extended Data Fig. 4c, d). In addition, a short helix ($\alpha 1$) was placed into the tube-like density adjacent to TM2, an assignment that was supported by the secondary structure search scores and sequence-based secondary structural prediction (Extended Data Fig. 4b). The final TARP model resembles a forearm and partially open right hand, with the TMD representing the

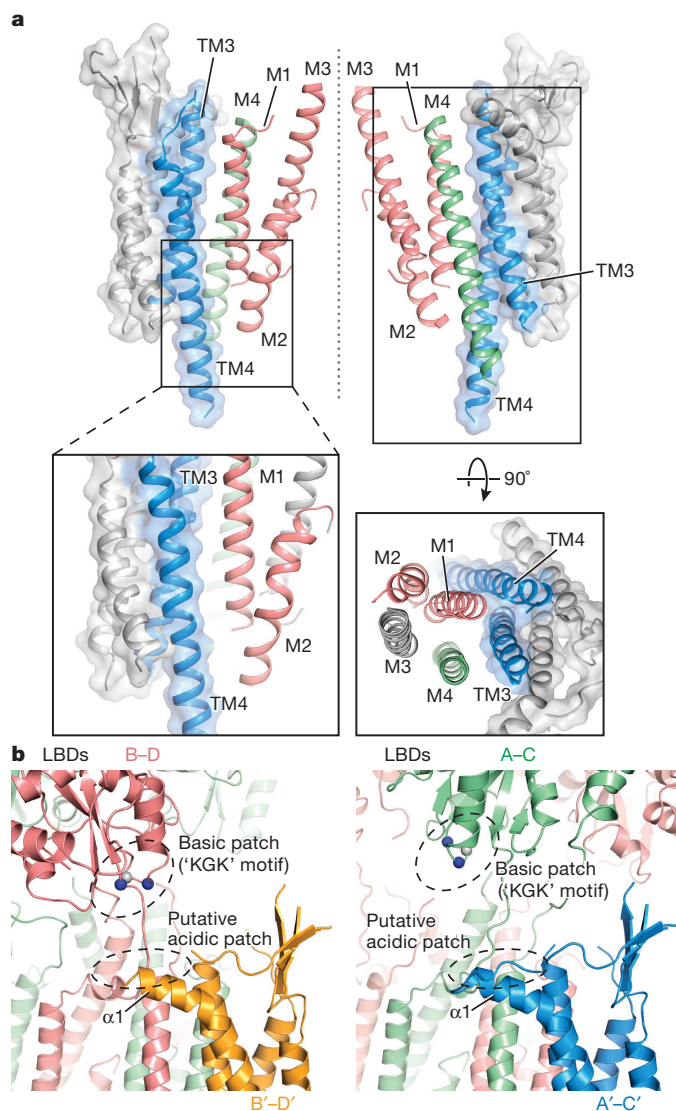


Figure 3 | Interactions between GluA2 receptor and TARP $\gamma 2$. **a**, TM–TM interactions between TARP $\gamma 2$ and receptor at A–C positions. TARP $\gamma 2$ is shown in ribbon diagram in transparent surface representation, whereas helices participating in interactions with receptor are highlighted in colour. The central axis of the ion channel pore is indicated by a dashed line. A close-up view emphasizes likely hydrophobic interactions between TARP and receptor, whereas a top-down view illustrates that all TM helices but M3 of the receptor interact with TARP. TM helices from selected receptor subunits were omitted for clarity. **b**, Interactions between TARP $\gamma 2$ ECD and receptor LBD at the A–C positions differ from interactions at the B–D positions. TARP $\gamma 2$ ECD and receptor LBD are in closer proximity in the B–D positions (right) than in the A–C positions (left). The C α atoms of the 'KGGK' motif (697–699) are shown as spheres. Lysine and glycine C α are coloured in blue and grey, respectively.

arm, the β -sheet representing the palm, and the short $\alpha 1$ helix representing the thumb (Extended Data Fig. 4c, d).

The closer proximity between LBDs and TARPs at the B–D positions compared to the A–C positions suggest that the A'–C' and B'–D' TARPs play non-equivalent roles in the modulation of receptor activities (Fig. 2c, d). Indeed, the lower lobes of the B–D LBDs have been proposed to play a greater role in ion channel gating²⁹. Nevertheless, it is possible that movements in the LBD layer upon agonist binding could cause the LBDs to engage the A'–C' TARPs, a hypothesis supported by evidence that binding of four TARPs leads to greater activation by the partial agonist kainate²⁵, highlighting the functional significance of TARP subunits in all four positions.

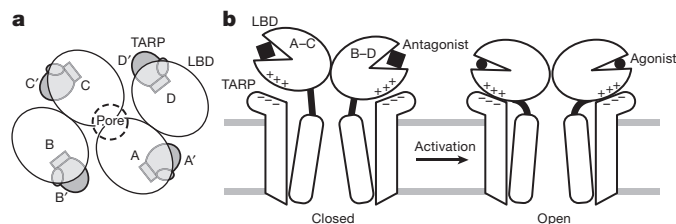


Figure 4 | Mechanism for TARP $\gamma 2$ modulation of receptor gating. **a**, TARP $\gamma 2$ subunits resemble partially opened palms and are positioned 'underneath' receptor LBDs in the antagonist-bound state. **b**, During receptor activation, TARP 'palms' engage with receptor LBD to stabilize intra-dimer and inter-dimer interfaces, modulating receptor activation, deactivation and desensitization. An extracellular loop of TARP $\gamma 2$ rich in negatively charged residues facilitates the motion of receptor LBD lower lobe rich in positively charged residues upon receptor activation, whereas TARP $\gamma 2$ TMD directly interacts with receptor TMD including the pore-lining M2 helices, leading to modulation of receptor pore properties.

The interactions between TARP and receptor are comprised of two components: TM–TM interactions and TARP extracellular domain (ECD)–receptor LBD interactions. The TM–TM interactions at both A–C and B–D positions are equivalent, obeying the four-fold symmetry of the TMD (Fig. 2e). The TM3 and TM4 of TARP form extensive hydrophobic interactions with M1 and M2 from one GluA2 subunit and with M4 from the adjacent subunit (Fig. 3a and Extended Data Fig. 6), suggesting one structural mechanism by which TARPs can modulate the properties of the ion channel. Given the nearly identical receptor TMD structure observed in apo and pre-open states, the TARP TMDs probably remain bound to receptor in a similar manner in activated states. By contrast, there are no direct contacts between the TARP thumb and palm and the receptor LBDs (Fig. 3b), although visualization of such interactions may be limited by the resolution of the reconstructions as well as inherent TARP flexibility. Nonetheless, a conserved acidic region spanning residues 85–95 (sequence: EDADYEADTAE) is present in the TARP extracellular 'loops' proximal to the $\alpha 1$ helix (Extended Data Fig. 4a), poised to interact with several positively charged residues on the lower lobe of the LBD, including the 'KGGK' sequence at residues 697–699 of the receptor³⁰ (Fig. 3b). Whereas these elements of structure may be too distant to form salt bridges in this antagonist-bound state, we speculate that such interactions could take place in pre-open or activated states (Extended Data Fig. 7), consistent with the importance of both TARP ECD and the LBD 'KGGK' motif, as well as a 'lowering' of the receptor LBD towards the membrane upon receptor activation^{12,13}.

Elucidation of the architecture of the GluA2 receptor–TARP $\gamma 2$ complex was facilitated by FSEC-based screening, which shows that digitonin stabilizes the receptor–TARP $\gamma 2$ complex. Analysis of the structure by single-particle cryo-EM illuminates how four TARPs encircle the receptor TMD and participate in extensive interactions with receptor TM helices, demonstrating the importance of non-polar contacts in complex formation. The acidic, partially open TARP 'palms' are positioned underneath basic motifs on the lower lobes of the LBDs (Fig. 4a), illustrating how complementary electrostatic interactions also contribute to receptor–TARP interactions (Fig. 4b). By juxtaposition of the TARP palms underneath the LBD clamshells, TARPs are ideally positioned to modulate domain closure and thus efficacy of partial agonists. Moreover, the A'–C' TARP proximity to the LBD dimer interface and the closeness of the B'–D' pair to the LBD dimer–dimer interface suggest how TARPs might modulate the activity of positive allosteric modulators and the modal gating properties of the receptor, respectively (Fig. 4a, b). We further speculate that the spatially distinct pairs of TARPs offer a structural explanation for biexponential kinetics deactivation and desensitization of the receptor–TARP $\gamma 2$ complex. Lastly, TARP TMD extensively interacts with receptor TMD including the pore helix M2, stabilizing the M2 helix and selectivity filter, thereby suggesting a mechanism for TARP modulation of receptor pore properties.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 May; accepted 24 June 2016.

Published online 1 July 2016.

1. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496 (2010).
2. Jackson, A. C. & Nicoll, R. A. The expanding social network of ionotropic glutamate receptors: TARPs and other transmembrane auxiliary subunits. *Neuron* **70**, 178–199 (2011).
3. Chen, L. *et al.* Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* **408**, 936–943 (2000).
4. Saitoh, Y. *et al.* Tight junctions. Structural insight into tight junction disassembly by *Clostridium perfringens* enterotoxin. *Science* **347**, 775–778 (2015).
5. Milstein, A. D. & Nicoll, R. A. Regulation of AMPA receptor gating and pharmacology by TARP auxiliary subunits. *Trends Pharmacol. Sci.* **29**, 333–339 (2008).
6. Shi, Y., Lu, W., Milstein, A. D. & Nicoll, R. A. The stoichiometry of AMPA receptors and TARPs varies by neuronal cell type. *Neuron* **62**, 633–640 (2009).
7. Tomita, S. *et al.* Stargazin modulates AMPA receptor gating and trafficking by distinct domains. *Nature* **435**, 1052–1058 (2005).
8. Milstein, A. D., Zhou, W., Karimzadeh, S., Bredt, D. S. & Nicoll, R. A. TARP subtypes differentially and dose-dependently control synaptic AMPA receptor gating. *Neuron* **55**, 905–918 (2007).
9. Soto, D., Coombs, I. D., Kelly, L., Farrant, M. & Cull-Candy, S. G. Stargazin attenuates intracellular polyamine block of calcium-permeable AMPA receptors. *Nat. Neurosci.* **10**, 1260–1267 (2007).
10. Nakagawa, T., Cheng, Y., Ramm, E., Sheng, M. & Walz, T. Structure and different conformational states of native AMPA receptor complexes. *Nature* **433**, 545–549 (2005).
11. Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756 (2009).
12. Chen, L., Dürr, K. L. & Gouaux, E. X-ray structures of AMPA receptor-cone snail toxin complexes illuminate activation mechanism. *Science* **345**, 1021–1026 (2014).
13. Dürr, K. L. *et al.* Structure and dynamics of AMPA receptor GluA2 in resting, pre-open, and desensitized states. *Cell* **158**, 778–792 (2014).
14. Meyerson, J. R. *et al.* Structural mechanism of glutamate receptor activation and desensitization. *Nature* **514**, 328–334 (2014).
15. Herguedas, B. *et al.* Structure and organization of heteromeric AMPA-type glutamate receptors. *Science* **352**, aad3873 (2016).
16. Kuusinen, A., Abele, R., Madden, D. R. & Keinänen, K. Oligomerization and ligand-binding properties of the ectodomain of the α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor subunit GluRD. *J. Biol. Chem.* **274**, 28937–28943 (1999).
17. Sun, Y. *et al.* Mechanism of glutamate receptor desensitization. *Nature* **417**, 245–253 (2002).
18. Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
19. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).
20. Armstrong, N., Jasti, J., Beich-Frandsen, M. & Gouaux, E. Measurement of conformational changes accompanying desensitization in an ionotropic glutamate receptor. *Cell* **127**, 85–97 (2006).
21. Keinänen, K. *et al.* A family of AMPA-selective glutamate receptors. *Science* **249**, 556–560 (1990).
22. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11–19 (1991).
23. Sommer, B. *et al.* Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* **249**, 1580–1585 (1990).
24. Letts, V. A. *et al.* The mouse stargazer gene encodes a neuronal Ca^{2+} -channel gamma subunit. *Nat. Genet.* **19**, 340–347 (1998).
25. Shanks, N. F., Maruo, T., Farina, A. N., Ellisman, M. H. & Nakagawa, T. Contribution of the global subunit structure and stargazin on the maturation of AMPA receptors. *J. Neurosci.* **30**, 2728–2740 (2010).
26. Turetsky, D., Garringer, E. & Patneau, D. K. Stargazin modulates native AMPA receptor functional properties by two distinct mechanisms. *J. Neurosci.* **25**, 7438–7448 (2005).
27. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
28. Turski, L. *et al.* ZK200775: a phosphonate quinoxalinedione AMPA antagonist for neuroprotection in stroke and trauma. *Proc. Natl Acad. Sci. USA* **95**, 10960–10965 (1998).
29. Kazi, R., Dai, J., Sweeney, C., Zhou, H. X. & Wollmuth, L. P. Mechanical coupling maintains the fidelity of NMDA receptor-mediated currents. *Nat. Neurosci.* **17**, 914–922 (2014).
30. Dawe, G. B. *et al.* Distinct structural pathways coordinate the activation of AMPA receptor-auxiliary subunit complexes. *Neuron* **89**, 1264–1276 (2016).

Acknowledgements We thank T. Nakagawa for providing the clone 10 cell line, Z. H. Yu, R. Huang and C. Hong (Janelia Campus) for assistance with microscope operation and data collection and the Advanced Computing Center (OHSU) for computational support. We are grateful to A. Goehring for help with cell culture, the Multiscale Microscopy Core (OHSU) for support with microscopy, L. Vaskalis for assistance with figures, H. Owen for help with proofreading and other Gouaux laboratory members for helpful discussions. S.C. is supported by an American Heart Association postdoctoral fellowship (16POST27790099). This work was supported by the NIH (E.G., NS038631). E.G. is an investigator with the Howard Hughes Medical Institute.

Author Contributions Y.Z., S.C. and E.G. designed the project. Y.Z. and S.C. performed sample preparation and cryo-EM data collection. Y.Z. and C.Y. analysed the data. I.B. performed electrophysiology experiments. Y.Z., S.C., C.Y. and E.G. wrote the manuscript with input from I.B.

Author Information The three-dimensional cryo-EM density map and the coordinates for the structure of the GluA2–TARP $\gamma 2$ complex have been deposited in the Electron Microscopy Data Bank and Protein Data Bank under the accession codes EMD-8256 and 5KK2, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G. (gouauxe@ohsu.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Electrophysiology. Electrophysiology experiments were performed using a stable cell line (clone 10) that constitutively expresses full-length wild-type TARP $\gamma 2$ and the C-terminally FLAG-tagged GluA2 AMPA receptor (flop isoform, arginine for the Q/R site) under control of the TerON promoter²⁵. Whole-cell recordings were carried out 10–18 h after induction of GluA2 receptor expression with $7.5 \mu\text{g ml}^{-1}$ doxycycline. Pipettes were pulled and polished to 2–3 M Ω resistance and filled with internal solution containing 75 mM CsCl, 75 mM CsF, 5 mM EGTA and 10 mM HEPES, pH 7.3. External solution contained 160 mM NaCl, 2.4 mM KCl, 4 mM CaCl₂, 4 mM MgCl₂, 10 mM HEPES pH 7.3 and $10 \mu\text{M}$ (R,R)-2b, a positive modulator that blocks desensitization³¹. To allow efficient binding of (R,R)-2b, each cell was perfused in external solution for one minute before currents were elicited by 3 mM of L-glutamate and 0.6 mM kainate, individually, with a one-minute wash step in between. Ratios of glutamate and kainate-evoked currents determined in five independent experiments were subjected to statistical analysis using Origin.

Expression and purification. The AMPA receptor–TARP $\gamma 2$ complex was expressed using clone 10 cells adapted to grow in suspension²⁵, using freestyle 293 expression medium supplemented with 2% (v/v) fetal bovine serum and selection antibiotics ($125 \mu\text{g ml}^{-1}$ zeocin, $150 \mu\text{g ml}^{-1}$ hygromycin, and $125 \mu\text{g ml}^{-1}$ neomycin). GluA2 (flop variant, arginine at the Q/R site) expression was induced by addition of $7.5 \mu\text{g ml}^{-1}$ doxycycline at a cell density of 2×10^6 cells per ml. Subsequently, MPQX²⁸ was added to the media (final concentration of 200 nM) to prevent cytotoxicity due to receptor overexpression. Cells were collected by centrifugation around 30–35 h after induction and homogenized by sonication. After removal of cell debris by centrifugation at 1,200g (15 min at 4 °C), the supernatant was subjected to ultracentrifugation at 100,000g for 1 h to collect the membrane fraction.

The membrane fraction was resuspended and solubilized in TBS buffer (20 mM Tris, pH 8.0, 150 mM NaCl) containing 1% (w/v) digitonin and $1 \mu\text{M}$ MPQX for 2 h at 4 °C. Insoluble material was removed by ultracentrifugation at 100,000g for 1 h, and the supernatant was passed through an anti-Flag immunoaffinity column pre-equilibrated with buffer P (20 mM Tris pH 8.0, 150 mM NaCl, 0.1% (w/v) digitonin, $1 \mu\text{M}$ MPQX), followed by a wash step using 10 column volumes of buffer P. The Flag-tagged GluA2 receptor in complex with TARP was eluted with buffer P supplemented with 0.5 mg ml^{-1} Flag peptide. The eluted complex was concentrated and further purified by size-exclusion chromatography (SEC) using a Superose 6 10/300 GL column equilibrated in Buffer P. Peak fractions were pooled and concentrated to 3 mg ml^{-1} using 100-kDa cut-off concentrator for subsequent biochemical analysis and cryo-EM studies.

Cryo-EM data acquisition. A droplet of $2.5 \mu\text{l}$ of purified AMPA receptor–TARP $\gamma 2$ complex at 3 mg ml^{-1} was placed on quantifoil 0.6–1.0 Au 200 mesh grids glow discharged at 30 mA for 120 s. The grid was then blotted for 2–3 s at 22 °C under conditions of 100% humidity, and flash-frozen in liquid ethane.

Cryo-EM data were collected on a 300 kV Titan Krios microscope using a K2 camera positioned post a GIF quantum energy filter. The energy filter was set to a 20 eV slit and a $70 \mu\text{m}$ objective aperture was used. Micrographs were recorded in super-resolution mode at a magnified physical pixel size of 1.35 Å, with the defocus ranging from -1.5 to $-2.5 \mu\text{m}$. Recorded at a dose rate of 8.3 e^- per pixel per s, each micrograph consisted of 40 dose-fractionated frames. Each frame was exposed for 0.3 s, resulting in a total exposure time of 12 s and total dose of 55 e^- per Å².

Image processing. A total of 2675 micrographs were subjected to motion correction with Unblur³². The CTF parameters for each micrograph were determined by CTFFIND3 (ref. 33) and particles were picked using DoG picker³⁴. Several rounds of 2D classification were used to remove ice contamination, micelles, disassociated or disordered protein and other false positives. The large number of particles discarded was likely a consequence of using DoGPickr with a fairly large threshold range; earlier attempts using template-based correlation manifested in an orientation bias during 2D classification. In this way, 2D classification also served as an opportunity to assess how well the selected particle orientations were distributed (Extended Data Fig. 3a). Rounds of 2D classification were repeated until the remaining classes had features recognizable from a comparison with an ensemble of 2D projections calculated by using the crystal structure of the antagonist-bound receptor. From an initial set of 257,378 putative particles, 61,539 particles were selected for subsequent 3D classification (Extended Data Fig. 2).

The subset of particles (61,539) was classified into four classes using a reference model generated from the GluA2 X-ray structure of the MPQX-bound state¹¹ (PDB code: 3KG2), which had been low-pass filtered to 60 Å. The most populated 3D class, containing 49% of total particles, featured four 4-fold symmetric

'bumps' on the extracellular side of the detergent micelle, and was subjected to further 3D refinement using a soft mask focused on the LBD and TMD domains³⁵, with C2 symmetry imposed³⁶. This further improved the quality of the density map, allowing the final map to reach 7.3 Å resolution as estimated by Fourier shell correlation between two independently refined half maps. Failing to show any putative TARP features and having poorly resolved features for the extracellular domains, the remaining three 3D classes possibly represented receptor free of TARP or even residual false positives, and were therefore excluded from subsequent analysis. All 2D and 3D classifications and refinements above were performed in RELION 1.4 (ref. 37).

Structural modelling. The structural modelling for the GluA2–TARP $\gamma 2$ complex was comprised of rigid-body fitting of LBDs and TMDs extracted from the MPQX-bound GluA2 crystal structure (PDB code: 3KG2) into the cryo-EM density, followed by fitting of a homologous TARP model generated by SWISS-MODEL³⁸ using the crystal structure of claudin-19 (ref. 4) (PDB code: 3X29) and sequence alignment performed with Clustal Omega³⁹. A 25.7% sequence similarity (11.3% identity) between claudin-19 and TARP $\gamma 2$ was determined by Sequence Manipulation Suite⁴⁰. Docked as a rigid body, the derived TARP homology model was refined in real space against the density map using COOT⁴¹ guided by the resolved helical density of the TARP TMs, and the density consistent with the conserved β -sheet on the extracellular side of the micelle. Furthermore, the TARP TM4 helical density was observed to protrude from the cytoplasmic side of the micelle, further assisting in the fitting of the TARP helices to the density map. The entire model was then improved by manual adjustments including removing several loop regions outside of density, local rigid-body fitting of individual helices into density, extension of the TARP TM4 helix by 14-residues and positioning of a short helix ($\alpha 1$) adjacent to TM2 supported by secondary structure prediction (Jpred4 (ref. 42)). The extension of the TARP TM4 helix was justified by strong density in the experimental density maps consistent with continuation of the α -helix, scoring in SSEhunter⁴³ consistent with an α -helix, and prediction of these residues in an α -helical conformation by secondary structure prediction (Jpred4 (ref. 42)).

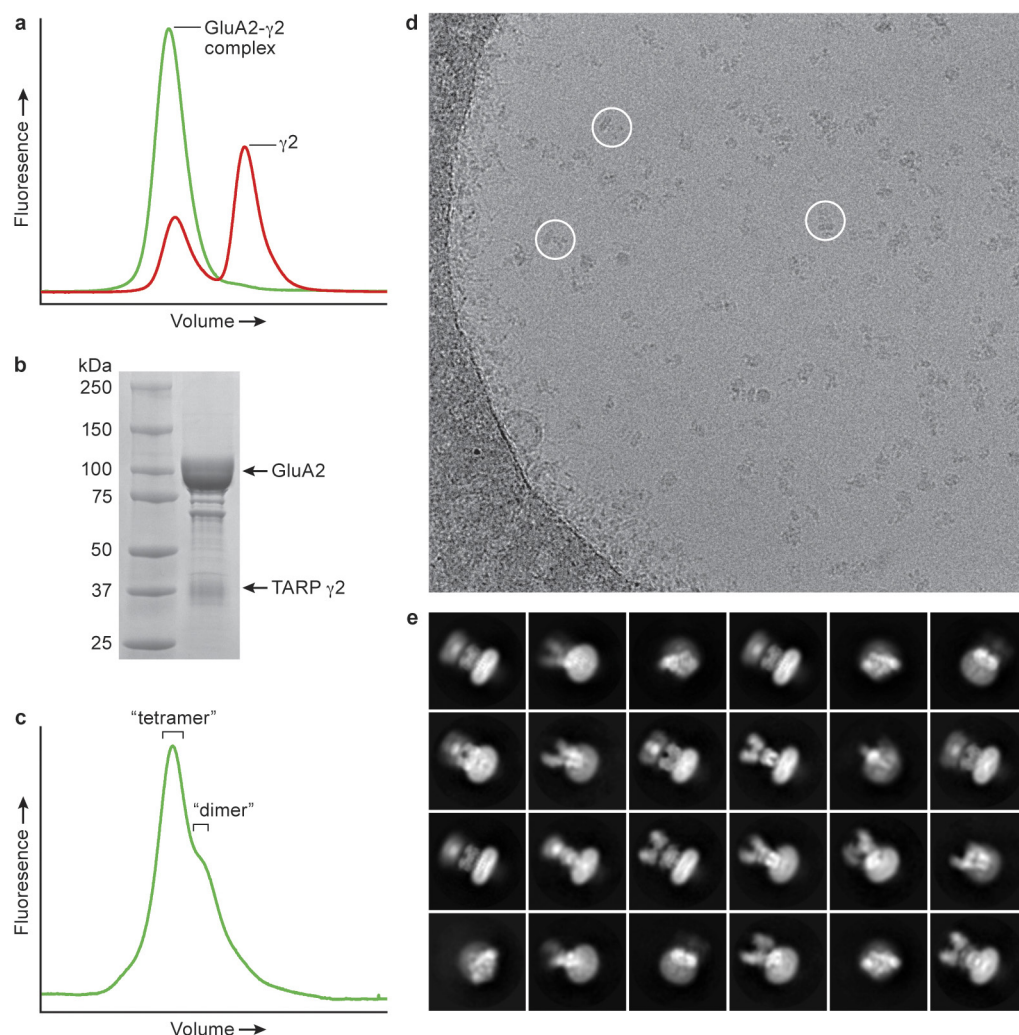
To validate the fitting and the placement of TARP TM4 extension and $\alpha 1$ helix, we used SSEhunter⁴³ to verify the secondary structure assignment against the EM density. To do this, the putative TARP density was extracted in Chimera⁴⁴ using Segger⁴⁵. SSEhunter analysis resulted in a series of pseudoatoms located on the skeleton of the density map, each assigned with a score. The positive scores at TM and $\alpha 1$ helix region and negative scores at the β -sheet region confirmed the secondary structure elements present in the TARP model.

The final map was put into a large P1 unit cell ($a = b = c = 405 \text{ Å}$; $\alpha = \beta = \gamma = 90^\circ$) and structure factors were calculated in PHENIX⁴⁶. The complex model of GluA2 receptor (LBD and TMD, residues 394–545, 564–587, 590–776, 781–826) and TARP (residues 6–38, 56–68, 72–82, (AC, 84–126)–(BD, 91–125), 131–162, 174–15; see also Extended Data Fig. 4) was then refined against structure factors derived from the density map using phenix.real_space_refine⁴⁶. Secondary structure, two-fold NCS and Ramachandran restraints were applied throughout the entire refinement. After refinement, map CC between model and EM map was 0.716, indicative of a reasonable fit at the present resolution. The resulting model was also used to calculate a model-map FSC curve, which agreed well with the gold-standard FSCs generated during the RELION refinement (Extended Data Fig. 3b). The final model has good stereochemistry, as evaluated using MolProbity (Extended Data Table 1).

All of the figures were prepared with Pymol⁴⁷, UCSF Chimera⁴⁴ and Prism 5.

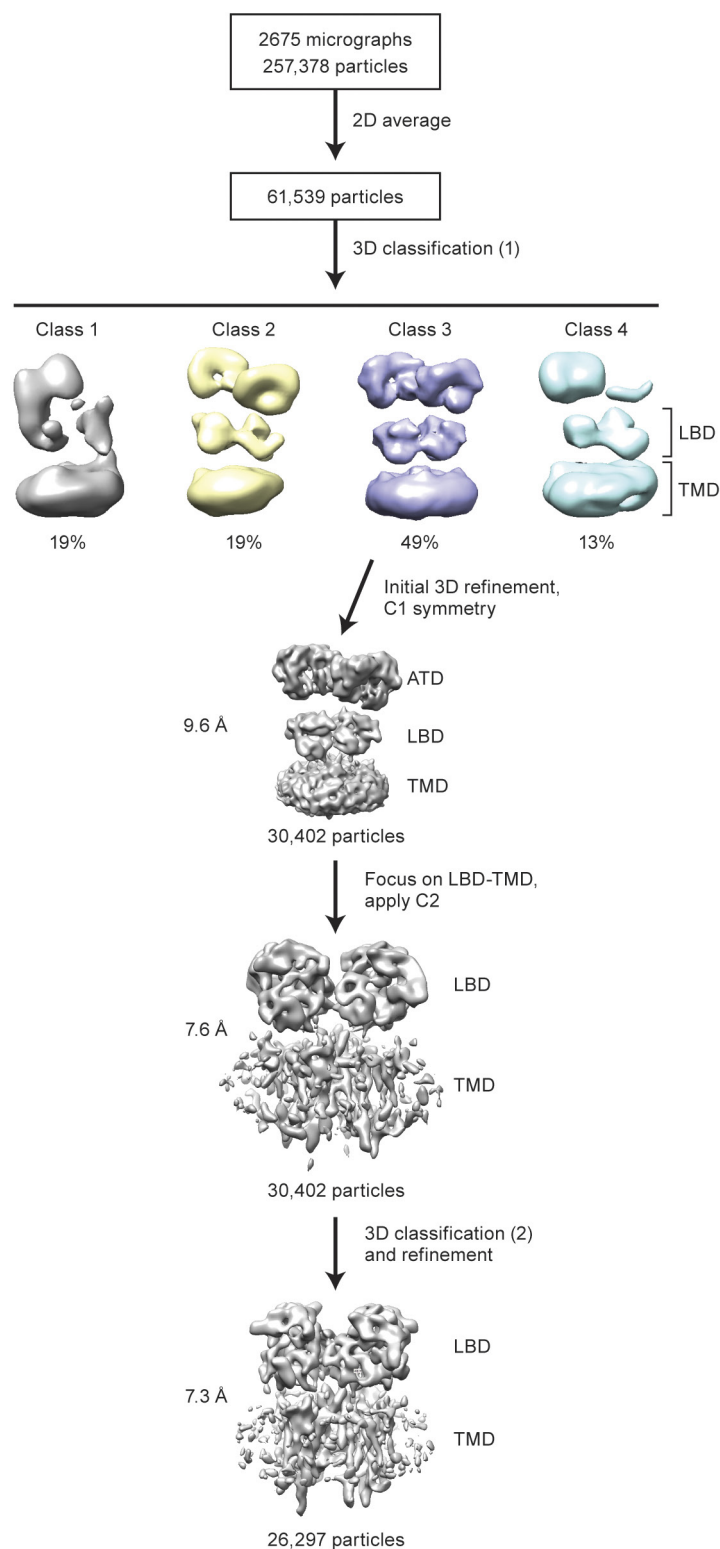
- Kaae, B. H. *et al.* Structural proof of a dimeric positive modulator bridging two identical AMPA receptor-binding sites. *Chem. Biol.* **14**, 1294–1303 (2007).
- Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Voss, N. R., Yoshioka, C. K., Radermacher, M. & Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).
- Penczek, P. A., Frank, J. & Spahn, C. M. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.* **154**, 184–194 (2006).
- Zhu, S. *et al.* Mechanism of NMDA receptor inhibition and activation. *Cell* **165**, 704–714 (2016).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
- Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

40. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**, 1102, 1104 (2000).
41. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
42. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43** (W1), W389–W394 (2015).
43. Baker, M. L., Ju, T. & Chiu, W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* **15**, 7–19 (2007).
44. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
45. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
46. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
47. The PyMOL Molecular Graphics System. (DeLano Scientific, San Carlos, California, USA, 2002).
48. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).



Extended Data Figure 1 | Digitonin is a suitable detergent for the purification of GluA2 receptor-TARP $\gamma 2$ complex. **a**, The receptor-TARP complex in digitonin disassociates when diluted into DDM. A complex composed of the GluA2 receptor and GFP-tagged TARP $\gamma 2$ was diluted in digitonin (green) or DDM (red) before being subjected to GFP-tuned FSEC analysis. **b**, Coomassie-blue-stained SDS-PAGE gel of the purified complex. **c**, Tryptophan-tuned FSEC profile of the purified complex was comprised of a major peak containing the tetrameric

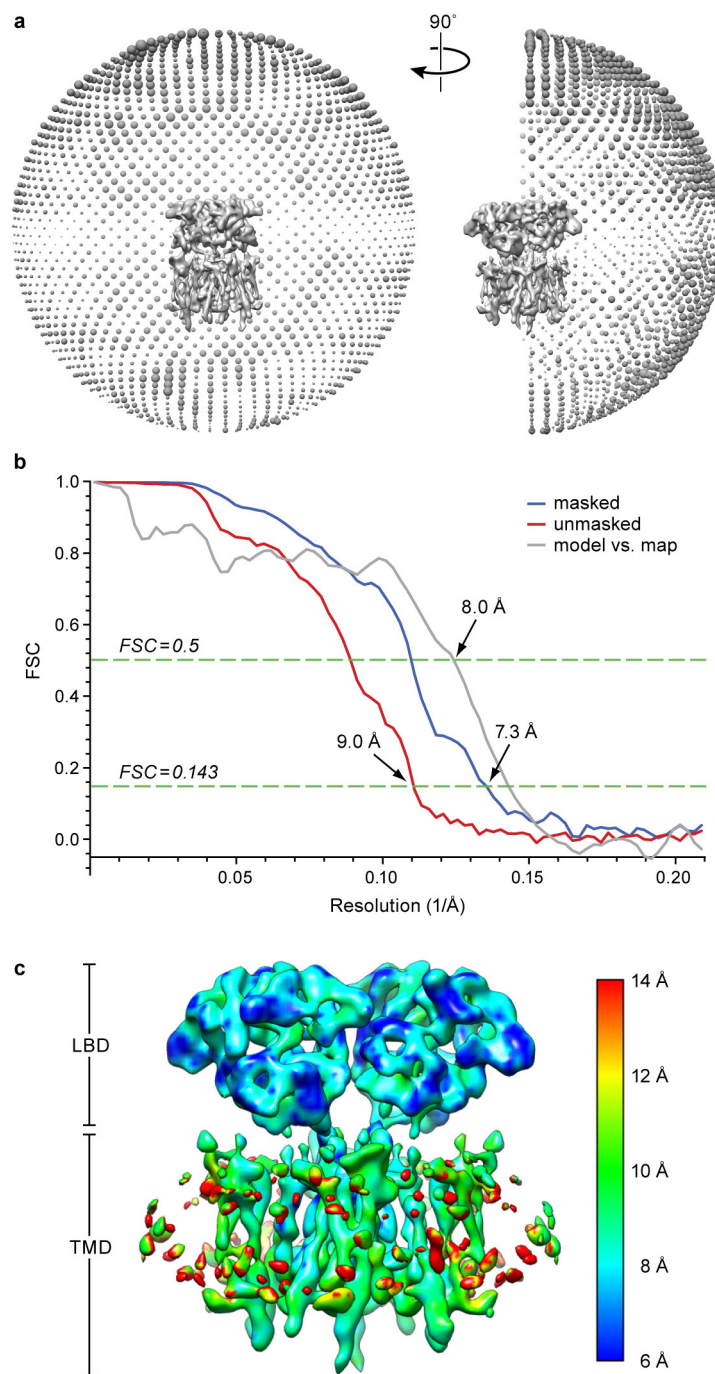
complex and a minor shoulder, the latter suggestive of either incompletely assembled or partially dissociated complexes. Only the full-size tetrameric species was used for single-particle cryo-EM analysis. **d**, A representative, motion-corrected micrograph of the GluA2 receptor-TARP $\gamma 2$ complex is shown. A few distinct complexes with the characteristic capital Y shape of the non-desensitized state of the AMPA receptor are circled. **e**, Representative 2D class averages showing a range of projections of the receptor-TARP $\gamma 2$ complex.



Extended Data Figure 2 | The work-flow of cryo-EM data processing.

The raw data set used in this study was composed of 2,675 micrographs. Particles (257,378) were picked from motion-corrected and contrast transfer function (CTF)-estimated micrographs for subsequent classifications. After multiple rounds of 2D classification, the remaining 61,539 particles were subjected to several rounds of 3D classification. Initial 3D classification yielded four major classes, where the most populated one contained 49% of total particles. An initial

3D reconstruction without imposed symmetry resulted in a moderate resolution at 9.6 Å. With C2 symmetry imposed, subsequent 3D refinement focused on the LBD and TMD layer improved the density map, allowing a reconstruction at 7.6 Å resolution. An additional two iterations of 3D classification and refinement using updated map as the reference further improved the reconstruction. The resolution of the final cryo-EM density map was estimated to be 7.3 Å.

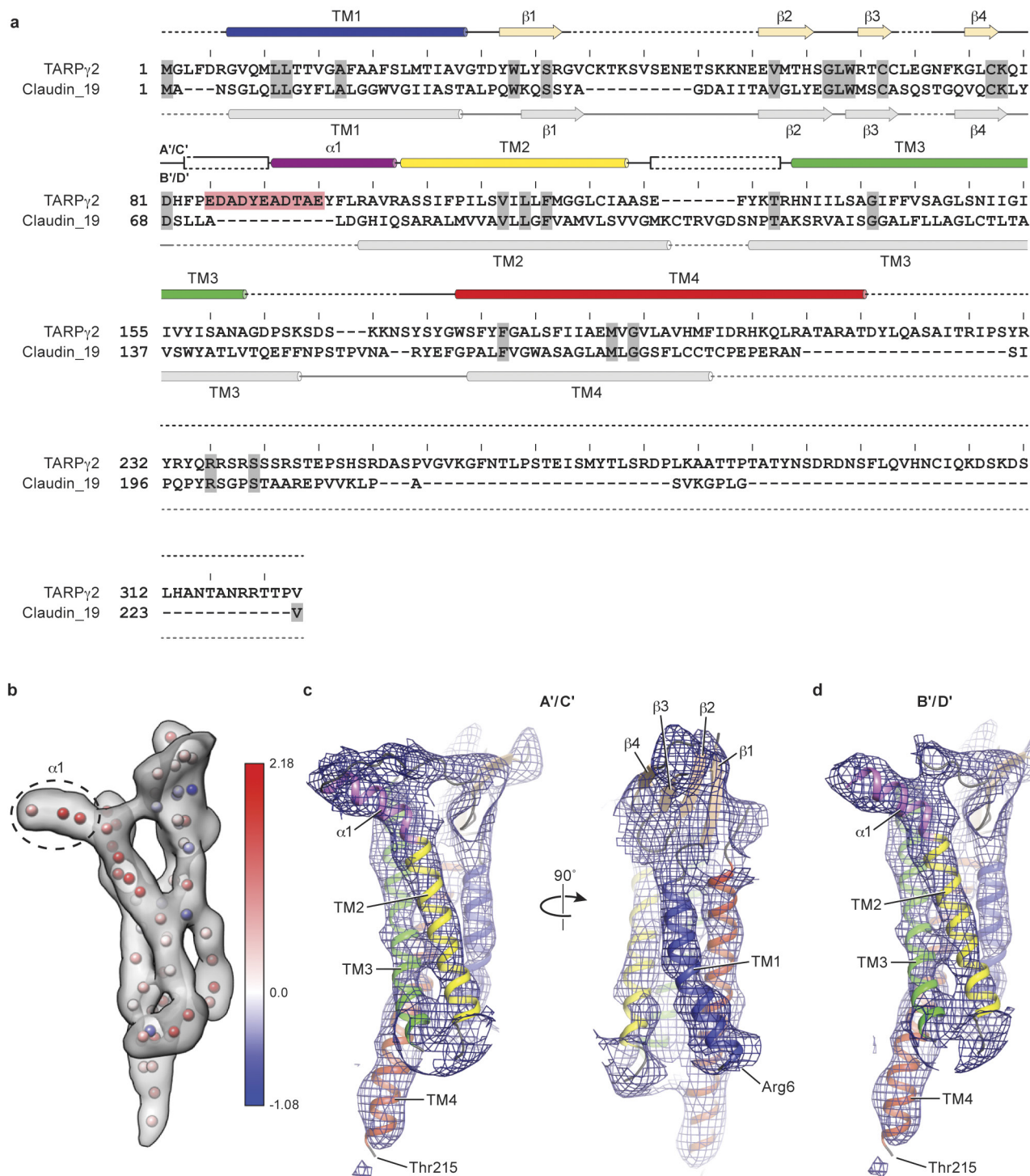


Extended Data Figure 3 | Statistics for the cryo-EM reconstruction.

a, Euler angular distribution of all particles included in the final 3D reconstruction. The number of particles viewed from each specific orientation was indicated by the size of the corresponding sphere.

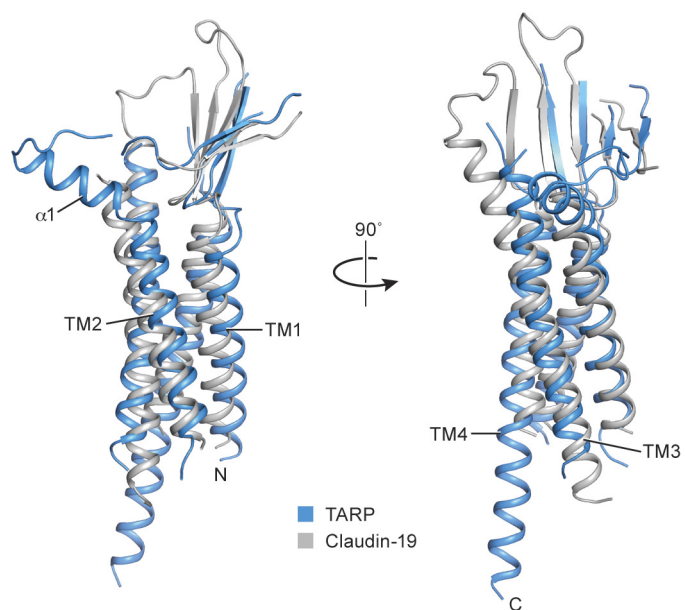
b, Gold-standard Fourier shell correlation (FSC) curves calculated

between two independently refined half-maps before (red) and after (blue) post-processing, overlaid with FSC curve calculated between cryo-EM density map and structural model. **c**, RESMAP⁴⁸ analysis of the unfiltered and unsharpened EM density map indicating the range of local resolution by colour code.

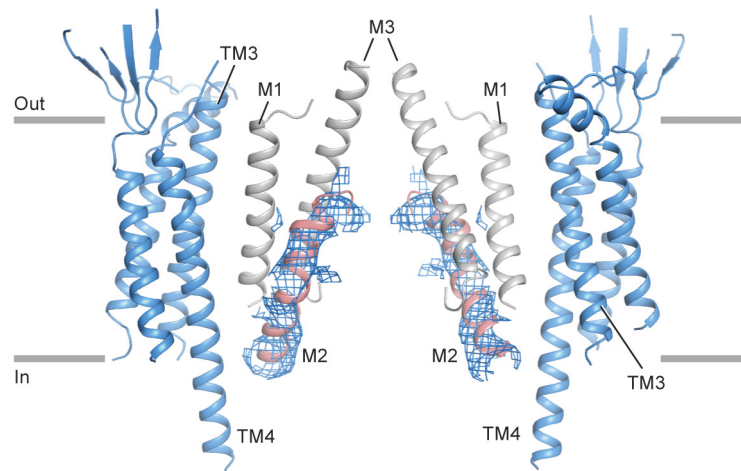


Extended Data Figure 4 | Structures of the TARP $\gamma 2$ subunits in the context of the respective cryo-EM density map. a, Sequence alignment between TARP $\gamma 2$ and claudin-19 calculated using Clustal omega. Also shown above the alignments are the secondary structure elements of TARP $\gamma 2$ based on the model reported here, and below the aligned sequences are the secondary structure elements derived from the claudin-19 crystal structure. The ECD region rich in negative charges is conserved throughout the TARP family and highlighted in red. **b**, EM density for B' TARP and pseudo-atoms placed by SSEhunter, each coloured according

to a calculated secondary structure score. Positive and negative scores indicate α -helix and β -sheet propensity, respectively. Dashed-line circles a map region where high scores were found, suggesting the presence of helical structure. A scale bar ranging from a maximum positive value (α -helix) to the minimum negative score (β -strand) is shown. **c**, The A' TARP of the A'-C' pair. **d**, The B' TARP of the B'-D' pair. The first and last visible residues Arg6 and Thr215 were labelled. Secondary structure elements were colour-coded as in a.

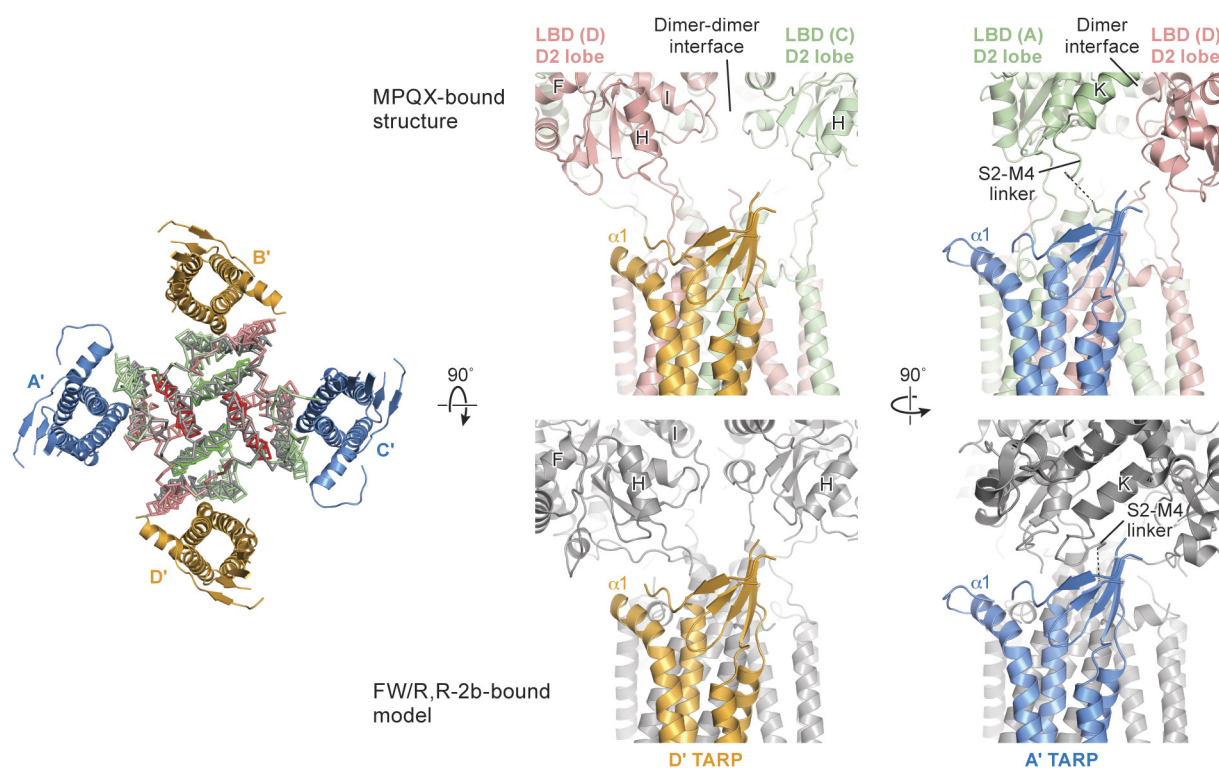


Extended Data Figure 5 | Structural comparison between TARP $\gamma 2$ and claudin-19. A superimposition of the TARP $\gamma 2$ structure (in blue) derived from this study and claudin-19 (in grey) is consistent in the conserved overall fold, with the exception that there is a short $\alpha 1$ helix present only in TARP $\gamma 2$.



Extended Data Figure 6 | Cryo-EM density map for the pore-helix of the GluA2 receptor. Clear density (blue mesh) is present for the pore-lining M2 helices, secondary structure elements that are weak or absent in all previous crystal structures. The N terminus of each pore helix is involved

in extensive interactions with TM4 from TARP subunits and we suggest that interactions of receptor TM helices that include M2, with TARP TM helices, stabilize the ion channel pore.



Extended Data Figure 7 | Possible interaction between receptor LBD and TARP in an active state. Shown on the left is a 'top-down view' of MPQX-bound receptor-TARP complex structure (in colour) superimposed with the crystal structure of an active state GluA2 receptor (in grey) in complex with a partial agonist, fluorowillardiine (FW) and a positive allosteric modulator, (R, R)-2b, using the central M3 helices as a reference. ATDs and LBDs were omitted for clarity. The modelled

pseudo-complex consisting of TARPs and the active state receptor illustrates a possible mechanism for how TARP interacts with receptor LBD during activation. Enlarged views of the MPQX-bound complex structure and FW-(R, R)-2b-bound complex model were shown side by side at both D' and A' TARP positions. LBD helices and a S2-M4 linker were labelled according to convention.

Extended Data Table 1 | Statistics of cryo-EM data collection, 3D reconstruction and model building

Data collection/processing	
Microscope	Krios
Voltage (kV)	300
Camera	Gatan K2
Camera mode	Super-resolution
Defocus range (μm)	-1.5 ~ -2.5
Exposure time (s)	12
Dose rate ($\text{e}^-/\text{pixel}/\text{s}$)	8.3
Magnified Pixel size (\AA)	1.35
Reconstruction	
Software	RELION 1.4
Symmetry	C2
Particles refined	26,297
Resolution (unmasked, \AA)	9.0
Resolution (masked, \AA)	7.3
Map sharpening B-factor (\AA^2)	-600
Model Statistics	
Protein residues	2309
Map CC	0.716
Resolution (FSC=0.5, \AA)	8.0
MolProbity score	2.49
C β deviations	0
Ramachandran	
Outliers	0.27%
Favored	88.28%
RMS deviations	
Bond length	0.008
Bond angles	1.26

RETRACTION

doi:10.1038/nature17648

Retraction: Effects of electron correlations on transport properties of iron at Earth's core conditions

Peng Zhang, R. E. Cohen & K. Haule

Nature **517**, 605–607 (2015); doi:10.1038/nature14090

In this Letter we reported density functional theory plus dynamical mean-field theory (DFT + DMFT) computations of the resistivity from electron–electron scattering at the conditions of Earth's core, and found that the electron–electron scattering was about the same magnitude as the conventional electron–phonon scattering, giving a total resistivity that was sufficient to allow a classical thermal-convection-driven dynamo. However, L. Pourovskii, J. Mravlje, S. Simak and I. Abrikosov could not reproduce our findings, which led us to re-examine our computations. We found an error of a factor of two that is due to our neglect of spin degeneracy (two electrons per band), which would halve the electron–electron resistivity and probably make the electron–electron scattering insignificant for the geodynamo, at least for pure iron. We therefore wish to retract this Letter.

The smaller electron–electron scattering supports the high conductivity of iron that was predicted¹ from electron–phonon density functional calculations. However, preliminary calculations show that using the exact double counting² recently developed for the DFT + DMFT method increases the electron–electron scattering. It is also probable that the Wiedemann–Franz law, assumed in our previous work, is not followed or has a non-constant Lorenz number in liquid metals³ or correlated systems⁴. Whether the resulting conductivity is consistent with a geodynamo driven by thermal convection requires further detailed calculations; the results will be reported elsewhere. The results and conclusions in the Letter that refer to resistivity at low temperatures (in Fig. 2b), and scattering rate and electronic structure (in Fig. 3) remain valid.

1. Sha, X. & Cohen, R. E. First-principles studies of electrical resistivity of iron under pressure. *J. Phys. Condens. Matter* **23**, 075401 (2011).
2. Haule, K. Exact double counting in combining the dynamical mean field theory and the density functional theory. *Phys. Rev. Lett.* **115**, 196403 (2015).
3. Yamasue, E., Susa, M., Fukuyama, H. & Nagata, K. Deviation from Wiedemann–Franz law for the thermal conductivity of liquid tin and lead at elevated temperature. *Int. J. Thermophys.* **24**, 713–730 (2003).
4. Joura, A. V., Demchenko, D. O. & Freericks, J. K. Thermal transport in the Falicov–Kimball model on a Bethe lattice. *Phys. Rev. B* **69**, 165105 (2004).

CORRIGENDUM

doi:10.1038/nature17984

Corrigendum: Structure of promoter-bound TFIID and model of human pre-initiation complex assembly

Robert K. Louder, Yuan He, José Ramón López-Blanco, Jie Fang, Pablo Chacón & Eva Nogales

Nature **531**, 604–609 (2016); doi:10.1038/nature17394

In this Article, author Pablo Chacón should have affiliation 4 (Department of Biological Physical Chemistry, Rocasolano Physical Chemistry Institute, CSIC, Serrano 119, Madrid 28006, Spain), not affiliation 3. This error has been corrected online.

CORRIGENDUM

doi:10.1038/nature17673

Corrigendum: Flexible high-temperature dielectric materials from polymer nanocomposites

Qi Li, Lei Chen, Matthew R. Gadinski, Shihai Zhang, Guangzu Zhang, Haoyu U. Li, Elissei Iagodkine, Aman Haque, Long-Qing Chen, Thomas N. Jackson & Qing Wang

Nature **523**, 576–579 (2015); doi:10.1038/nature14647

The author Elissei Iagodkine was erroneously omitted from the author list of this Letter. They are associated with the affiliation: Dow Chemical Company, 455 Forest Street, Marlborough, Massachusetts 01752, USA, and the Author Contributions section should have included the statement: E.I. provided research-grade BCB used in the preparation of the samples reported, and also participated in helpful discussions. This has been corrected in the online versions of the paper. In addition, authors Haoyu Li and Tom Jackson should have been listed as Haoyo U. Li and Thomas N. Jackson, respectively; the author list, Acknowledgements and Author Contributions sections have therefore been altered accordingly.

CAREERS

TECHNOLOGY IN CHECK The right brakes for gene drive **p.117**

ROYAL CONVERSATION Countess hosts meeting for Nobel laureates **go.nature.com/2a1y0d2**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

RANDSC/ALAMY STOCK PHOTO



A good conference poster can get you noticed and spur discussions that can expand your network.

CONFERENCE PRESENTATIONS

Lead the poster parade

An eye-catching presentation can attract potential collaborators — and even a cash prize.

BY CHRIS WOOLSTON

Stroll through a decent-sized scientific conference and you'll probably face a bewildering number of posters — many more than you could ever hope to read in one day. So you have to pick and choose. Perhaps one reminds you of the worst PowerPoint presentation you ever endured; another is crammed with thousands of words in microscopic font. But you encounter one featuring a bold illustration, splashes of colour, readable text and clean lines. You pause for a closer look, chat with the presenter and discover common research interests. You've made a connection, and at least one poster has accomplished what its creator meant it to do.

The scientific poster remains a crucial

currency for communication and connection, says biophysicist Anthony Salvagno, director of education for #SciFund Challenge, a non-profit organization in Santa Barbara, California, that specializes in science-communication training. Through SciFund, he co-teaches a five-week online course on poster design along with biologist Zen Faulkes of the University of Texas Rio Grande Valley in Edinburg.

Researchers now have access to an array of high-end graphics software — and the 'how to make a poster' conversation has been going on for years (see *Nature* **483**, 113–115; 2012). But that hasn't stemmed the flow of visual clunkers. As Salvagno explains, researchers often slap posters together at the last minute instead of thinking about the best ways to deliver their message and engage their audience.

But those who have the vision — and computer skills — to avoid distracting design blunders will draw the right kind of attention to themselves, their findings and their ideas. They might even win an award (see 'Tips for making your poster stand out'), although the main goals are to publicize their science and scientific identity while forging new associations. "A good poster will help you make better connections," Salvagno says. "Just one conversation can turn into a huge success."

Trishna Dutta, a wildlife researcher at Columbia University in New York City who studies tigers in India, says that lessons from the poster course helped to spark productive conversations at the 2015 International Congress for Conservation Biology (ICCB) in Montpellier, France. She had signed up for the course specifically ►

► to make an impression at the conference. She also wanted to make up for past failures. “My first posters were bad,” she says. “I didn’t have the aesthetic sense of what goes with what.” Worse, comments from attendees suggested that her key points were often lost, especially for those outside her speciality. “That was a case where I needed to know my audience,” she says. “People there studied everything from bacteria to elephants. I’m not sure they got my message.”

Her ICCB poster was far clearer. A subheading spelled out the take-away message of tiger migration, the text was orderly and easy to read, maps added colour as well as context, and a photo of a wild tiger near the centre captured the eye. “I still don’t make excellent posters, but I’m getting the hang of it,” she says.

Anxiety about these visual presentations is widespread. When Vasco Elbrecht uploaded a set of scientific-poster tutorials on YouTube (go.nature.com/2akrsly), he realized that he had underestimated the demand for such help. “I would have been happy if just a few of my friends watched them,” says the PhD student at the University of Duisburg-Essen in Germany. So far, his poster tutorials have racked up more than 31,900 views.

In his most-viewed video, Elbrecht shows examples of good and bad posters from his own repertoire. His first — about the genetics of *Microbotryum* fungus — was bogged down with huge swathes of text, a common pitfall. “I tried to fit everything I could on it,” he says. “But at a conference, nobody is going to stand there

and read it for ten minutes.” In a later, more successful poster about the genetic diversity of the stonefly *Dinocras cephalotes*, he limited the text to a few hundred words — roughly the same count as an abstract (see ‘A winning view’). That’s generally enough to deliver a key message and entice passers-by without overwhelming them, he says. The design also helped him to win a €1,500 (US\$1,660) research prize for his poster and abstract from the Institute for the Advancement of Water Quality and Water Resources Management in Essen, Germany, in 2014.

LESS IS MORE

Salvagno and Faulke’s poster class stresses the same point: when it comes to text, less is more. Poster-makers often already know that too much text can be off-putting, but many are still unable to resist the temptation to include practically everything they know about their subject. “When I ask people what they dislike about posters, too much text is the number-one complaint,” Salvagno says. “People hate seeing it on other people’s posters, but they do it on their own.”

Of course, there’s more to it than getting the right word count. Text and graphics have to flow together in a way that’s as visually appealing as it is informative. That takes a designer’s eye — or a willingness to copy from people who know what they are doing. Elbrecht encourages researchers to borrow elements from posters that they like. “All design is redesign,” he says. “There’s no need to be original.”

Effective posters take many shapes, but they tend to have some basic elements in common, says Sam Hertig, a freelance scientific illustrator in Berne, Switzerland. Hertig, who has just completed a postdoc in computational biology, gave a talk on creating a visually striking scientific poster at Stanford University in California earlier this year and uploaded the presentation to YouTube (go.nature.com/2aetlrc). As he explains, a “stunning” poster generally starts with a gripping centrepiece image, whether of a molecule, organism or galaxy. One of his own recent posters featured a multicoloured image of HIV. “Be daring,” he says in the presentation. “There may be hundreds or thousands of posters at a conference. You want something that will stand out.”

Hertig says that the text of a poster should have its own visual appeal. In most cases, the text will be neatly arranged in 2 to 4 columns on a poster that’s about 91 cm by 122 cm. The font, which should be consistent throughout, must be clear and easy to read (not something like Comic Sans), and should be at least 24 points.

The poster should be printed to the maximum size allowed by the conference, and the title should be large and legible from a distance. The subheadings — which should also be clear and visible — should say something more dynamic than ‘Results’. If, for instance, research uncovered a 5% decline in the reproductive success of heat-stressed frogs, the heading for the results section should hint at that finding.

Hertig says that the placement of white

EYES ON THE PRIZE

Tips for making your poster stand out

Good posters are supposed to communicate results and foster connections — but a first-place ribbon wouldn’t hurt, either. A poster prize is more than a badge of honour: it’s an accomplishment that would look great on a CV. Here are some tips for getting the prize.

- **Scientific modelling.** Winning posters often go beyond flat text and graphics. Where appropriate, consider building a 3D model of your study subject. “It doesn’t add any scientific value, but it gets people’s attention,” says Vasco Elbrecht, a PhD student at the University of Duisburg-Essen in Germany who won a cash prize for his poster in 2014.

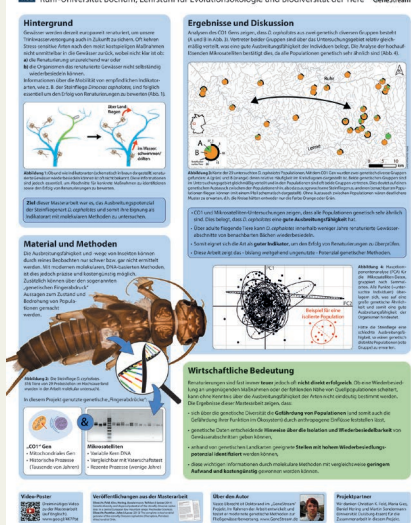
- **Tech it up.** Technology has opened up new possibilities: some conferences allow attendees to bolster their posters with videos on a tablet or similar device. “If you really want to go for a poster prize, have a QR smartphone barcode for a video on your topic,” Elbrecht says.

- **Do a test run.** Before you ever set foot in a conference, you should be confident that your poster has all the clarity, appeal and impact that you intended. “The important thing is to get honest feedback,” Elbrecht says. “Show it to people in another department if necessary.”

- **Know your audience.** Hedwig van der Meer, a PhD student at the Amsterdam University of Applied Sciences in the Netherlands, was an underdog at the 2016 American Academy of Orofacial Pain poster session in Florida. “I was a physical therapist from the Netherlands going up against all of these American doctors,” she says. “I didn’t think I could win, especially after seeing the other posters.” Her presentation was heavy with text but short on colour. Yet it worked because the presentation and the topic — the connection between temporomandibular disorders and headaches — hit the sweet spot. “The audience was a match,” she says. “I had a clear message, and I’m passionate about what I do.” **C.W.**

Insekten als Indikatoren für Renaturierungserfolge

Vasco Elbrecht (Vasco.Elbrecht@uni-due.de), Ralph Tullrian & Florian Leese
Ruhr-Universität Bochum, Lehrstuhl für Evolutionsökologie und Biodiversität der Tiere



A WINNING VIEW

Vasco Elbrecht’s award-winning poster about the genetic diversity of the stonefly *Dinocras cephalotes* has an eye-catching centrepiece image and limits text to a few hundred words — enough to deliver a key message and entice passers-by.

space is an important but often overlooked aspect of poster design. Visually attractive posters tend to have substantial borders and significant gaps between text blocks. The white space should flow together in a cohesive way that draws in the eye while giving it a chance to rest. In a room full of posters screaming for attention, he says, some well-placed emptiness can offer tranquility.

THE RIGHT TOOL FOR THE JOB

Yet these design aesthetics won't amount to much without the right software. Many researchers resort to PowerPoint, usually because they already have PowerPoint figures at hand. It can work: Hedwig van der Meer, a physiotherapy PhD student at the Amsterdam University of Applied Sciences in the Netherlands, used PowerPoint to make her first-place poster at the 2016 conference of the American Academy of Orofacial Pain in Orlando, Florida. But Salvagno advises against the program: it isn't designed for printing, the colours may be off and the alignment tools are cumbersome. If PowerPoint is the only option, he recommends disabling the 'snap to grid' function for maximum control of the layout.

Hertig recommends vector-based graphics programs such as Inkscape or Adobe Illustrator. Unlike PowerPoint and other programs that create illustrations with pixels, both of these use equations to determine each point; images and text can thus be scaled up without loss of clarity. These programs can also smoothly align text and captions. Choose one vector-based program and stick with it for every poster and presentation, Hertig adds. "It's important to invest the time early in your PhD. You won't have to learn it again. It will just be natural."

A quality poster is just one part of a successful presentation. At most conferences, the presenter will have at least a couple of hours to stand by their posters and interact with attendees. This is where some of the most important work at a conference takes place, which is why researchers should spend as much time polishing their pitches as they spend creating their poster, Salvagno says. He recommends preparing several different versions of one's talking points: a 20-second elevator pitch for the mildly curious and a longer version for anyone who wants a deeper dive.

For her part, van der Meer thinks that her presentation of her prizewinning poster was as important as the actual product. "You have to involve the audience by being open and enthusiastic," she says. "The combination of a clear poster and passionate presentation works best, because people will understand your work and get excited." ■

Chris Woolston is a freelance writer in Billings, Montana.

TURNING POINT

Kevin Esvelt

Evolutionary engineer Kevin Esvelt, at the Massachusetts Institute of Technology in Cambridge, works with gene drives, engineered bits of DNA that can cause a mutation to become heritable all the time. He calls for researchers to create and use safe lab procedures while working with this powerful but potentially risky technology.

What is a gene drive?

In nature, a gene drive occurs when a DNA sequence spreads through a population by breaking the conventional rules of inheritance. For example, if an organism has a single copy of a fluorescent marker gene and its mate has none, normally only half their offspring will fluoresce. When a gene-drive system is in play, almost all of them will glow.

How can scientists use this capability?

Gene drives allow us to drive altered traits through wild populations over generations. For instance, we could alter the DNA of wild mosquitoes to stop them from carrying disease. We could restore damaged ecosystems and save endangered wildlife by genetically removing invasive species.

How did your insights help to propel this field?

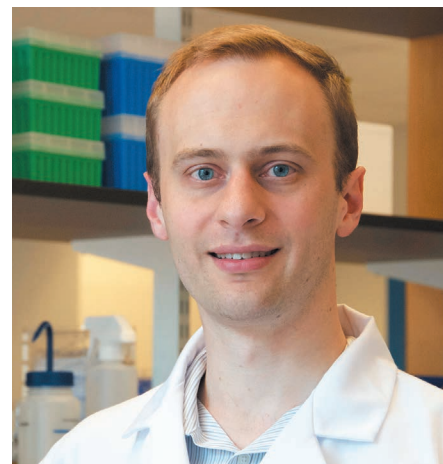
Even ten years ago, heritable genome editing was a possibility, but no one had found a molecular tool that would enable it to be done efficiently. In 2013, laboratories began using CRISPR to precisely edit the genomes of many species. I realized then that this tool could be used to build stable gene drives in many complex organisms. It could also be used to build reverse drives, which are like molecular erasers for overwriting previous edits.

Why did you explain how gene drives would work before you published results showing that they could work in any organism?

Most advances don't give individual scientists the power to affect entire ecosystems. By detailing what was possible, how it could be achieved and what safeguards were needed to prevent any accidental release of altered organisms from the lab, we hoped to set an example of how future work in gene drives should proceed.

Why was this important?

A single escaped organism that found a mate could eventually alter most of the local population and, very possibly, every population of that species worldwide. The ecological risk might be low, but the damage to public trust in biotechnology could imperil the future of the field.



Did you want researchers to agree on some guidelines first?

My immediate priority was to prevent the accidental release of any gene-drive organisms into the wild. I wrote to the few researchers working on gene drives to explain my concerns about ethics and safety.

What happened?

Last year, we released results showing that gene drives work in yeast. Then another group — who were working with fruit flies — independently created a functional gene-drive system. They were careful to keep the flies contained, but unlike our paper, their manuscript, which was meant to be published as a how-to for other labs, made no mention of safeguards or the risk to wild populations. To their credit, they agreed to include those details.

Did your efforts help to usher in regulation?

The fruit-fly case triggered responses from many scientists. For months, we struggled to agree on which safeguards should be used in the lab. We eventually published our recommendations in July 2015, and this year the US National Academy of Sciences released a report setting out how to conduct gene-drive research responsibly.

Should gene-drive information be classified?

Classifying such information would hinder beneficial applications and threaten biosecurity. We must know which species to monitor. Open science is the best defence and the best way to earn public support. ■

INTERVIEW BY VIJEE VENKATRAMAN

This interview has been edited for length and clarity.

FLOATING IN MY TIN CAN

Lullaby for life.

BY GERRI LEEN

I always thought that if I died in space, I would take my last breaths surrounded by stars and the detritus of the fleet of the five federations. But I've wandered off course, my new nav system is failing, and now I'm running blind with a life-support system that is dying.

I have enemies. They know how to sabotage a ship. I check: I always check. But the new nav system's readings must have masked whatever they did to the life support. It waited to fail until I was too far from anywhere to call for help.

No beacons ping on my comm system, not even from the farthest planets out.

Genina, obsessive and damaged, was from one of the outer planets. Her voice was lovely when she sang. I wish I had a recording of her, especially when she sang the lullaby her mother had sung to her.

Her voice made me feel safe, the way I hadn't since I was a child. Before men from the first federation came and took my parents away. Before I was handed over to protective services and sent to the children's home, where the refuse of the five federations go.

Where they taught freedom as a code word for conforming. Where we had what was expected drilled into us. But at night, when the lights were out, Genina and I fought back in the room that we shared. We discovered things that made us unique and incandescent, like the stars I longed to fly among. Genina never wanted space; she wanted to be free to sing her songs.

My dreams were of steering a ship anywhere that would make us free.

In the children's home, Genina would sing quietly as we shared her bed, skin to skin, the covers pulled over us, until one night the matrons found out and separated us.

After that, I'd see Genina in the halls, and when she spoke, her voice was different, cracked and dusty, as if she'd been denied water.

It was hard to get her alone, but I finally did. She wouldn't touch me and she didn't want to sing. But I begged and eventually she sang the lullaby. I wept — what came out of her mouth was no longer song.

She didn't cry; she just touched my cheek and went back to her room. They found her floating in the river a week later. I think it was my fault — if I hadn't asked her to sing,



she might not have realized what she'd lost.

I leave the ship on autopilot and check the cryo units, normally well hidden, but there's no reason to keep them camouflaged now that I'm floating blind. My cargo is asleep, deep in frozen dreams. They paid me to get them out of the five federations. They are singers, you see.

I rescue singers from those who would destroy the songs.

Or I did. Now ... now I drift with them safely in cryo and think about ways to make this right.

I suppose I should console myself with knowing I *did* get them free of the five federations. I just didn't get them all the way. Would that be a consolation? That they're free for now?

They sang for me, these two lovers who wanted to share a cryo pod until I explained how that wouldn't work. The way they looked at each other reminded me of how Genina used to smile at me. Their voices trilled in half-step harmonies that made me shiver. So discordant and yet ... beautiful.

These two should have been in their new home by now, being lauded by audiences, after they'd sung one last song for me — I asked all my runners for that. Credits, too, of course — a song wouldn't buy fuel — but that last song always sounded the sweetest. Hope coloured it with something beautiful.

But now it's hopeless and we float and wait for our air to run out. It will be a while, from my perspective. No

time at all as far as the sleepers know.

Would they want to spend the time left to them singing? Or just dreaming of song?

I finger the console. An almost musical combination of keystrokes will wake them up.

Or I can let them sleep — the cold changing their slumber to a more lasting rest as the power runs out, while the inside of the ship begins to resemble their cryo chamber and I freeze to death.

But ... I long to hear their voices again. Maybe they know Genina's lullaby. I didn't ask when they boarded. They're not from her planet, but songs travel.

I push down on the first key. With a ping, 'Initiate awakening?' blinks on the screen.

I have done this before, five times. Each time I have chosen 'No'.

I hit 'No' again.

I try to remember how Genina's song went. My voice cracks and bends around notes that may be right but certainly sound nothing like those she made so magical.

"I'm sorry," I say to Genina, as I say every night.

As I say now to these two lovers, whom I have failed.

As I say to myself, as I turn back to the command console, take my seat, and try to figure a way out of this even though I know there isn't one.

I think of the cascading trill of my two frozen lovebirds. Wouldn't they rather die together, knowing their time was ending? Singing?

Or is that just what I want?

Every time I go back to the cryo chambers, I ask myself this, and I let them sleep.

Genina's lullaby would sound like a dirge in two-step harmony. It would be ... fitting.

No — I will let them sleep.

Won't I? ■

Gerri Leen lives in Northern Virginia and originally hails from Seattle. She has stories and poems published by Daily Science Fiction, Escape Pod, Grimdark, Enchanted Conversation and others. Her first solo editing gig, *the A Quiet Shelter* There anthology published by Hadley Rille Books, was released in autumn 2015 and benefits homeless animals. See more at <http://www.gerrileen.com>.

ILLUSTRATION BY JACEY